AN ACCELERATED LINEARIZED ALTERNATING DIRECTION METHOD OF MULTIPLIERS

YUYUAN OUYANG*, YUNMEI CHEN†, GUANGHUI LAN‡, AND EDUARDO PASILIAO JR. §

Abstract. We present a novel framework, namely AADMM, for acceleration of linearized alternating direction method of multipliers (ADMM). The basic idea of AADMM is to incorporate a multi-step acceleration scheme into linearized ADMM. We demonstrate that for solving a class of convex composite optimization with linear constraints, the rate of convergence of AADMM is better than that of linearized ADMM, in terms of their dependence on the Lipschitz constant of the smooth component. Moreover, AADMM is capable to deal with the situation when the feasible region is unbounded, as long as the corresponding saddle point problem has a solution. A backtracking algorithm is also proposed for practical performance.

1. Introduction. Assume that $\mathcal{W}$, $\mathcal{X}$ and $\mathcal{Y}$ are finite dimensional vectorial spaces equipped with inner product $\langle\cdot,\cdot\rangle$, norm $\|\cdot\|$ and conjugate norm $\|\cdot\|_*$. Our problem of interest is the following affine equality constrained composite optimization (AECCO) problem:

$$\min_{x\in X, w\in\mathcal{W}} G(x)+F(w),\ \text{s.t.}\ Bw-Kx=b, \tag{1.1}$$

where $X\subseteq\mathcal{X}$ is a closed convex set, $G(\cdot):X\to\mathbb{R}$ and $F(\cdot):\mathcal{W}\to\mathbb{R}$ are finitely valued, convex and lower semi-continuous functions, and $K:X\to\mathcal{Y}$, $B:\mathcal{W}\to\mathcal{Y}$ are bounded linear operators.

In this paper, we assume that $F(\cdot)$ is simple, in the sense that the optimization problem

$$\min_{w\in\mathcal{W}} \frac{\eta}{2}\|w-c\|^2+F(w),\ \text{where } c\in\mathcal{W},\eta\in\mathbb{R} \tag{1.2}$$

can be solved efficiently. We will use the term "simple" in this sense throughout this paper, and use the term "non-simple" in the opposite sense. We assume that $G(\cdot)$ is non-simple, continuously differentiable, and that there exists $L_G>0$ such that

$$G(x_2)-G(x_1)-\langle\nabla G(x_1),x_2-x_1\rangle\le\frac{L_G}{2}\|x_2-x_1\|^2,\ \forall x_1\in X,x_2\in X. \tag{1.3}$$

One special case of the AECCO problem in (1.1) is when $B=I$ and $b=0$. Under this situation, problem (1.1) is equivalent to the following unconstrained composite optimization (UCO) problem:

$$\min_{x\in X} f(x):=G(x)+F(Kx). \tag{1.4}$$

Both AECCO and UCO can be reformulated as saddle point problems. By the method of Lagrangian multipliers, the AECCO problem (1.1) is equivalent to the following saddle point problem:

$$\min_{x\in X,w\in\mathcal{W}}\max_{y\in\mathcal{Y}} G(x)+F(w)-\langle y,Bw-Kx-b\rangle. \tag{1.5}$$

The AECCO and UCO problems have found numerous applications in machine learning and image processing. In most application, $G(\cdot)$ is known as the fidelity term and $F(\cdot)$ is the regularization term. For example, consider the following two dimensional total variation (TV) based image reconstruction problem

$$\min_{x\in\mathbb{F}^n}\frac{1}{2}\|Ax-c\|^2+\lambda\|Dx\|_{2,1}, \tag{1.6}$$

*Department of Industrial and System Engineering, University of Florida (ouyang@ufl.edu). Part of the research was done while the author was a PhD student at the Department of Mathematics, University of Florida. This author was partially supported by AFRL Mathematical Modeling Optimization Institute.

†Department of Mathematics, University of Florida (yun@math.ufl.edu). This author was partially supported by NSF grants DMS-1115568, IIP-1237814 and DMS-1319050.

‡Department of Industrial and System Engineering, University of Florida (glan@ise.ufl.edu). This author was partially supported by NSF grant CMMI-1000347, ONR grant N00014-13-1-0036, NSF DMS-1319050, and NSF CAREER Award CMMI-1254446..

§Munitions Directorate, Air Force Research Laboratory (pasiliao@eglin.af.mil)

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
FEB 2014		**00-00-2014 to 00-00-2014**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
An Accelerated Linearized Alternating Direction Method of Multipliers	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of California Los Angeles,Department of Mathematics,520 Portola Plaza,Los Angeles,CA,90095**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

We present a novel framework, namely AADMM, for acceleration of linearized alternating direction method of multipliers (ADMM). The basic idea of AADMM is to incorporate a multi-step acceleration scheme into linearized ADMM. We demonstrate that for solving a class of convex composite optimization with linear constraints, the rate of convergence of AADMM is better than that of linearized ADMM, in terms of their dependence on the Lipschitz constant of the smooth component. Moreover, AADMM is capable to deal with the situation when the feasible region is unbounded, as long as the corresponding saddle point problem has a solution. A backtracking algorithm is also proposed for practical performance.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**32**	

where the field $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$, x is the n-vector form of a two-dimensional complex or real valued image, $D : \mathbb{F}^n \to \mathbb{F}^{2n}$ is the two-dimensional finite difference operator acting on the image x, and

$$\|y\|_{2,1} := \sum_{i=1}^{n} \|(y^{(2i-1)}, y^{(2i)})^T\|_2, \ \forall y \in \mathbb{F}^{2n},$$

where $\|\cdot\|_2$ is the Euclidean norm in $\mathbb{R}^2$. In (1.6), the regularization term $\|Dx\|_{2,1}$ is the discrete form of TV semi-norm. By setting $G(x) := \|Ax - c\|^2/2$, $F(\cdot) := \|\cdot\|_{2,1}$, $K = \lambda D$, $X = \mathcal{X} = \mathbb{F}^n$ and $\mathcal{W} = \mathbb{F}^{2n}$, problem (1.6) becomes a UCO problem in (1.4).

1.1. Notations and terminologies. In this subsection, we describe some necessary assumptions, notations and terminologies that will be used throughout this paper.

We assume that there exists an optimal solution (w^*, x^*) of (1.1) and that there exists $y^* \in \mathcal{Y}$ such that $z^* := (w^*, x^*, y^*) \in \mathcal{Z}$ is a saddle point of (1.5), where $\mathcal{Z} := \mathcal{W} \times \mathcal{X} \times \mathcal{Y}$. We also use the notation $Z := \mathcal{W} \times X \times Y$ if a set $Y \subseteq \mathcal{Y}$ is declared readily. We use $f^* := G(x^*) + F(w^*)$ to denote the optimal objective value of problem (1.1). Since UCO problems (1.4) are special cases of AECCO (1.1), we will also use f^* to denote the optimal value $G(x^*) + F(Kx^*)$.

In view of (1.1), both the objective function value and the feasibility of the constraint should be considered when defining approximate solutions of AECCO, henceforth the following definition comes naturally:

DEFINITION 1.1. *A pair $(w, x) \in \mathcal{W} \times X$ is called an (ε, δ)-solution of (1.1) if*

$$G(x) + F(w) - f^* \le \varepsilon, \ \text{and} \ \|Bw - Kx - b\| \le \delta.$$

We say that (w, x) has primal residual ε and feasibility residual δ. In particular, if (w, x) is an $(\varepsilon, 0)$-solution, then we simply say that it is an ε-solution.

The feasibility residual δ in Definition 1.1 measures the violation of the equality constraint, and the primal residual ε measures the gap between the objective value $G(x) + F(w)$ at the approximate solution and the optimal value f^*. For an (ε, δ)-solution (w, x) where $\delta > 0$, since (w, x) does not satisfy the equality constraint in (1.1), it is possible that $G(x) + F(w) - f^* < 0$. However, as pointed out in [31], a lower bound of $G(x) + F(w) - f^*$ is given by

$$G(x) + F(w) - f^* \ge \langle y^*, Bw - Kx - b\rangle \ge -\delta\|y^*\|,$$

where y^* is a component of $z^* = (w^*, x^*, y^*)$, a saddle point of (1.5).

In the remainder of this subsection, we introduce some notations that will be used throughout this paper. The following distance constants will be used for simplicity:

$$\begin{aligned} &D_{w^*,B} := \|B(w_1 - w^*)\|, D_{x^*,K} := \|K(x_1 - x^*)\|, D_{x^*} := \|x_1 - x^*\|, D_{y^*} := \|y_1 - y^*\|, \\ &D_{X,K} := \sup_{x_1,x_2\in X} \|Kx_1 - Kx_2\|, \ \text{and} \ D_S := \sup_{s_1,s_2\in S} \|s_1 - s_2\|, \ \text{for any compact set } S. \end{aligned} \tag{1.7}$$

For example, for any compact set $Y \subset \mathcal{Y}$, we use D_Y to denote the diameter of Y. In addition, we use $x_{[t]}$ to denote sequence $\{x_i\}_{i=1}^t$, where x_i's may either be real numbers, or points in vectorial spaces. We will also equip a few operations on the notation of sequences. Firstly, suppose that $\mathcal{V}_1$, $\mathcal{V}_2$ are any vector spaces, $v_{[t+1]} \subset \mathcal{V}_1$ is any sequence in $\mathcal{V}_1$ and $\mathcal{A} : \mathcal{V}_1 \to \mathcal{V}_2$ is any operator, we use $\mathcal{A}v_{[t+1]}$ to denote the sequence $\{\mathcal{A}v_i\}_{i=1}^{t+1}$. Secondly, if $\eta_{[t]}, \tau_{[t]} \subset \mathbb{R}$ are any real valued sequences, and $L \in \mathbb{R}$ is any real number, then $\eta_{[t]} - L\tau_{[t]}$ denotes $\{\eta_i - L\tau_i\}_{i=1}^t$. Finally, we denote by $\eta_{[t]}^{-1}$ the reciprocal sequence $\{\eta_i^{-1}\}_{i=1}^t$ for any non-zero real valued sequence $\eta_{[t]}$.

1.2. Augmented Lagrangian and alternating direction method of multipliers. In this paper, we study AECCO problems from the aspect of the augmented Lagrangian formulation of (1.5):

$$\min_{x\in X, w\in\mathcal{W}} \max_{y\in\mathcal{Y}} G(x) + F(w) - \langle y, Bw - Kx - b\rangle + \frac{\rho}{2}\|Bw - Kx - b\|^2, \tag{1.8}$$

where ρ is a penalty parameter. The idea of analyzing (1.8) in order to solve (1.1) is essentially the augmented Lagrangian method (ALM) by Hestenes [26] and Powell [44] (It is originally called the method of multipliers in [26, 44]; see also the textbooks, e.g., [5, 41, 6]). The ALM is a special case of the Douglas-Rachford splitting method [19, 16, 32], which is also an instance of the proximal point algorithm [17, 46]. The iteration complexity of an inexact version of ALM, where the subproblems are solved iteratively by Nesterov's method, has been studied in [30]. One influential variant of ALM is the ADMM algorithm [20, 21], which is an alternating method for solving (1.8) by minimizing x and w alternatively and then updating the Lagrangian coefficient y (See [7] for a comprehensive explanation on ALM, ADMM and other algorithms). In compressive sensing and imaging science, the class of Bregman iterative methods is an application of the ALM and the ADMM. In particular, the Bregman iterative method [24] is equivalent to ALM, and the split Bregman method [23] is equivalent to ADMM.

We give a brief review on ADMM, and some of its variants. The scheme of ADMM is described in Algorithm 1.

Algorithm 1 The alternating direction method of multipliers (ADMM) for solving (1.1)

Choose $x_1 \in X$, $w_1 \in \mathcal{W}$ and $y_1 \in \mathcal{Y}$.
for $t = 1, \ldots, N-1$ **do**

$$x_{t+1} = \operatorname*{argmin}_{x \in X} G(x) - \langle y_t, Bw_t - Kx - b\rangle + \frac{\rho}{2}\|Bw_t - Kx - b\|^2, \tag{1.9}$$

$$w_{t+1} = \operatorname*{argmin}_{w \in \mathcal{W}} F(w) - \langle y_t, Bw - Kx_{t+1} - b\rangle + \frac{\rho}{2}\|Bw - Kx_{t+1} - b\|^2, \tag{1.10}$$

$$y_{t+1} = y_t - \rho(Bw_{t+1} - Kx_{t+1} - b). \tag{1.11}$$

end for

For non-simple G, a linearized ADMM (L-ADMM) scheme generates iterate x_{t+1} in (1.9) by

$$x_{t+1} = \operatorname*{argmin}_{x \in X} \langle \nabla G(x_t), x\rangle + \langle y_t, Kx\rangle + \frac{\rho}{2}\|Bw_t - Kx - b\|^2 + \frac{\eta}{2}\|x - x_t\|^2. \tag{1.12}$$

We may also linearize $\|Bw_t - Kx - b\|^2$, and generate x_{t+1} by

$$x_{t+1} = \operatorname*{argmin}_{x \in X} G(x) + \langle y_t, Kx\rangle - \rho\langle Bw_t - Kx_t - b, Kx\rangle + \frac{\eta}{2}\|x - x_t\|^2, \tag{1.13}$$

as discussed in [18, 10]. This variant is called the preconditioned ADMM (P-ADMM). If we linearize both $G(x)$ and $\|Bw_t - Kx - b\|^2$, we have the linearized preconditioned ADMM (LP-ADMM), in which (1.9) is changed to

$$x_{t+1} = \operatorname*{argmin}_{x \in X} \langle \nabla G(x_t), x\rangle + \langle y_t, Kx\rangle - \rho\langle Bw_t - Kx_t - b, Ax\rangle + \frac{\eta}{2}\|x - x_t\|^2. \tag{1.14}$$

There has been several works on the convergence analysis and applications of ADMM, L-ADMM, and P-ADMM. It is shown in [10] that P-ADMM (Algorithm 1 with $\theta = 1$ in [10]) solves the UCO problem with rate of convergence

$$\mathcal{O}\left(\frac{\|K\|D^2}{N}\right),$$

where N is the number of iterations and D depends on the distances D_{x^*} and D_{y^*}. There are also several works concerning the tuning of the stepsize η_t in L-ADMM, including [50, 51, 11].

For AECCO problems, in [34] ADMM is treated as an instance of block-decomposition hybrid proximal extragradient (BD-HPE), and it is proved that the rate of convergence of the primal residual of ADMM for solving AECCO is

$$\mathcal{O}\left(\frac{D^2}{N}\right),$$

where D depends on B, D_{x^*} and D_{y^*}. In [25], the convergence analysis of ADMM and P-ADMM is studied based on the variational inequality formulation of (1.5), in which similar rate of convergence is achieved under the assumption that both the primal and dual feasible sets in (1.5) are bounded. In [42], it is shown that if X is compact, then the rate of convergence of ADMM and L-ADMM for solving the AECCO problem is

$$G(x^N) + F(w^N) - f^* + \rho\|Bw^N - Kx^N - b\|^2 \le \mathcal{O}\left(\frac{L_G D_X^2 + \rho D_{y^*,B}^2}{N}\right), \forall \rho > 0, \tag{1.15}$$

where (x^N, w^N) is the average of iterates $x_{[N]}$ of the ADMM algorithm. The result in (1.15) is stronger than the results in [34, 25], in the sense that both primal and feasibility residuals are included in (1.15), while in [34, 25] there is no discussion on the feasibility residual. However, the rate of convergence of the feasibility residual is still not very clear in (1.15), considering that $G(x^N) + F(w^N) - f^*$ can be negative.

1.3. Accelerated methods for AECCO and UCO problems. In a seminal paper [39], Nesterov introduced a smoothing technique and a fast first-order method that solves a class of composite optimization. When applied to UCO problems, Nesterov's method has optimal rate of convergence

$$\mathcal{O}\left(\frac{L_G D_{x^*}^2}{N^2} + \frac{\|K\| D_{x^*} D_Y}{N}\right)^{1}, \tag{1.16}$$

where Y is the bounded dual space of the UCO problem. Following the breakthrough in [40], much effort has been devoted to the development of more efficient first-order methods for non-smooth optimization (see, e.g., [38, 1, 29, 15, 43, 48, 4, 28]). Although the rate in (1.16) is also $\mathcal{O}(1/N)$, what makes it more attractive is that it allows very large Lipschitz constant L_G. In particular, L_G can be as large as $\Omega(N)$, without affecting the rate of convergence (up to a constant factor). However, it should be noted that the boundedness of Y is critical for the convergence analysis of Nesterov's smoothing scheme. Following [40], there has also been several studies on the AECCO and UCO problems, and it has been shown that better acceleration results can be obtained if more assumptions are enforced for the AECCO and UCO problem. We give a list of such assumptions and results.

1). *Excessive gap technique.* The excessive gap technique is proposed in [38] for solving the UCO problem in which G is simple. Comparing to [40], the method in [38] does not require the total number of iterations N to be fixed in advance. Furthermore, if $G(\cdot)$ is strongly convex, it is shown that the rate of convergence of the excessive gap technique is $\mathcal{O}(1/N^2)$.

2). *Special instance.* For the UCO problem, if $K = I$ and G is simple, an accelerated method with skipping steps is proposed in Algorithm 7 of [22], which achieves $\mathcal{O}(1/N^2)$ rate of convergence. The result is better than (1.16), but with cost of evaluating objective value functions in each iteration. For AECCO problem with compact feasible sets, it is shown in [33] that if $G(\cdot)$ is a composition of a strictly convex function and a linear transformation and $F(\cdot)$ is the weighted sum of 1-norm and some 2-norms, the asymptotic rate of convergence of ADMM method and its variants is R-linear.

3). *Strong convexity.* In [10] for solving the UCO problem in which G is simple, the authors showed that P-ADMM is equivalent to their proposed method, and furthermore, if either $G(\cdot)$ or $F^*(\cdot)$ is uniformly convex, then the rate of convergence of their method can be accelerated to $\mathcal{O}(1/N^2)$. It is worth noting that this rate of convergence is weaker since it uses a different termination criterion.

[1] It is assumed in [40] that X is compact, hence the rate of convergence is dependent on D_X. However, the analysis in [40] is also applicable for the case when X is unbounded, yielding (1.16).

In addition, if both $G(\cdot)$ and $F^*(\cdot)$ are uniformly convex (hence the objective function in (1.4) is continuously differentiable), the proposed method in [10] converges linearly. When both $G(x)$ and $F(x)$ are strongly convex in the AECCO problem, an accelerated ADMM method is proposed in [23], which achieves the $\mathcal{O}(1/N^2)$ rate of convergence.

It should be noted that all the methods in the above list require more assumptions on the AECCO and UCO problems (e.g., simplicity of $G(\cdot)$, strong convexity of $G(\cdot)$ or $F(\cdot)$), in comparison with Nesterov's smoothing scheme. More recently, we proposed an accelerated primal-dual (APD) method for solving the UCO problem [13], which has the same optimal rate of convergence (1.16) as that of Nesterov's smoothing scheme in [40]. The advantage of the APD method over Nesterov's smoothing scheme is that it does not require boundedness on either X or Y. The basic idea of the APD method is to incorporate a multi-step acceleration into LP-ADMM, and this has motivated our studies on accelerating the linearized ADMM method for solving the AECCO and UCO problems.

1.4. Contribution of the paper. The main interest of this paper is to develop an accelerated linearized ADMM algorithm for solving AECCO and UCO problems, in which G is a general convex and non-simple function. Our contribution in this paper mainly consists of the following aspects.

Firstly, we propose an accelerated framework for ADMM (AADMM), which consists two novel accelerated linearized ADMM methods, namely, accelerated L-ADMM (AL-ADMM) and accelerated LP-ADMM (ALP-ADMM). We prove that AL-ADMM and ALP-ADMM have better rates of convergence than L-ADMM and LP-ADMM in terms of their dependence on L_G. In particular, we prove that both accelerated methods can achieve rates similar to (1.16), hence both of them can efficiently solve problems with large Lipschitz constant L_G (as large as $\Omega(N)$). We show that L-ADMM and LP-ADMM are special instances of AL-ADMM and ALP-ADMM respectively, with rates of convergence $\mathcal{O}(1/N)$. To improve the performance in practice, we also propose a simple backtracking technique for searching Lipschitz constants L_G and $\|K\|$.

Secondly, the proposed framework solve both AECCO and UCO problems with unbounded feasible sets, as long as a saddle point of problem (1.5) exists. Instead of using the perturbation type gap function in [13], our convergence analysis is performed directly on both the primal and feasibility residuals. The estimate of the rate of convergence will depend on the distance from the initial point to the set of optimal solutions.

2. An accelerated ADMM framework. In this section, we propose an accelerated ADMM framework for solving AECCO (1.1) and UCO (1.4). The proposed framework, namely AADMM, is presented in Algorithm 2.

In AADMM, the binary constant χ in (2.2) is either 0 or 1, the superscript "ag" stands for "aggregate", and "md" stands for "middle". It can be seen that the middle point x_t^{md}, and the aggregate points w_{t+1}^{ag}, x_{t+1}^{ag} and y_{t+1}^{ag} are weighted sums of all the previous iterates $\{x_i\}_{i=1}^t$, $\{w_i\}_{i=1}^{t+1}$, $\{x_i\}_{i=1}^{t+1}$ and $\{y_i\}_{i=1}^{t+1}$, respectively. If the weights $\alpha_t \equiv 1$, then $x_t^{md} = x_t$ and the aggregate points are exactly the current iterates w_{t+1}, x_{t+1} and y_{t+1}. In this case, if $\chi = 0$, and $\theta_t = \tau_t = \rho_t \equiv \rho$, then AADMM becomes L-ADMM, and if in addition G is simple, then AADMM becomes ADMM. On the other hand, if $\chi = 1$, then AADMM becomes LP-ADMM, and if in addition G is simple, AADMM becomes P-ADMM.

In this work, we will show that if G is non-simple, by properly specifying the parameter α_t, we can significantly improve the rate of convergence of Algorithm 2 in terms of its dependence on L_G, with about the same iteration cost. We call the acceleration for $\chi = 0$ the accelerated L-ADMM (AL-ADMM), and call that for $\chi = 1$ the accelerated LP-ADMM (ALP-ADMM).

Next, we define certain appropriate gap functions.

2.1. Gap functions. For any $\tilde{z} = (\tilde{w}, \tilde{x}, \tilde{y}) \in \mathcal{Z}$ and $z = (w, x, y) \in \mathcal{Z}$, we define

$$Q(\tilde{w}, \tilde{x}, \tilde{y}; w, x, y) := [G(x) + F(w) - \langle \tilde{y}, Bw - Kx - b\rangle] - [G(\tilde{x}) + F(\tilde{w}) - \langle y, B\tilde{w} - K\tilde{x} - b\rangle]. \tag{2.8}$$

For simplicity, we use the notation $Q(\tilde{z}; z) := Q(\tilde{w}, \tilde{x}, \tilde{y}; w, x, y)$, and under different situations, we may use notations $Q(\tilde{z}; w, x, y)$ or $Q(\tilde{w}, \tilde{x}, \tilde{y}; z)$ for the same meaning. We can see that $Q(z^*, z) \geq 0$ and $Q(z, z^*) \leq 0$ for all $z \in \mathcal{Z}$, where z^* is a saddle point of (1.5), as defined in Section 1.1. For compact sets $W \subset \mathcal{W}, X \subset$

Algorithm 2 Accelerated ADMM (AADMM) framework

Choose $x_1 \in X$ and $w_1 \in \mathcal{W}$ such that $Bw_1 = Kx_1 + b$. Choose Set $x_1^{ag} = x_1$, $w_1^{ag} = w_1$ and $y_1^{ag} = y_1 = 0$.
for $t = 1, \ldots, N-1$ **do**

$$x_t^{md} = (1-\alpha_t)x_t^{ag} + \alpha_t x_t, \tag{2.1}$$

$$x_{t+1} = \operatorname*{argmin}_{x\in X} \langle \nabla G(x_t^{md}), x\rangle - \chi\theta_t\langle Bw_t - Kx_t - b, Kx\rangle + \frac{(1-\chi)\theta_t}{2}\|Bw_t - Kx - b\|^2 + \langle y_t, Kx\rangle + \frac{\eta_t}{2}\|x - x_t\|^2, \tag{2.2}$$

$$x_{t+1}^{ag} = (1-\alpha_t)x_t^{ag} + \alpha_t x_{t+1}, \tag{2.3}$$

$$w_{t+1} = \operatorname*{argmin}_{w\in\mathcal{W}} F(w) - \langle y_t, Bw\rangle + \frac{\tau_t}{2}\|Bw - Kx_{t+1} - b\|^2, \tag{2.4}$$

$$w_{t+1}^{ag} = (1-\alpha_t)w_t^{ag} + \alpha_t w_{t+1}, \tag{2.5}$$

$$y_{t+1} = y_t - \rho_t(Bw_{t+1} - Kx_{t+1} - b), \tag{2.6}$$

$$y_{t+1}^{ag} = (1-\alpha_t)y_t^{ag} + \alpha_t y_{t+1}. \tag{2.7}$$

end for
Output $z_N^{ag} = (w_N^{ag}, x_N^{ag})$.

$\mathcal{X}, Y \subset \mathcal{Y}$, the duality gap function

$$\sup_{\tilde{w}\in W, \tilde{x}\in X, \tilde{y}\in Y} Q(\tilde{w}, \tilde{x}, \tilde{y}; w, x, y) \tag{2.9}$$

measures the accuracy of an approximate solution (w, x, y) to the saddle point problem

$$\min_{x\in X, w\in W}\max_{y\in Y} G(x) + F(w) - \langle y, Bw - Kx - b\rangle.$$

However, our problem of interest (1.1) has a saddle point formulation (1.5), in which the feasible set $(\mathcal{W}, X, \mathcal{Y})$ may be unbounded. Recently, a perturbation-based termination criterion is employed by Monteiro and Svaiter [35, 36, 34] for solving variational inequalities and saddle point problems. This termination criterion is based on the enlargement of a maximal monotone operator, which is first introduced in [8]. One advantage of using this termination criterion is that its definition does not depend on the boundedness of the domain of the operator. We modify this termination criterion and propose a modified version of the gap function in (2.9). More specifically, we define

$$g_Y(v, z) := \sup_{\tilde{y}\in Y} Q(w^*, x^*, \tilde{y}; z) + \langle v, \tilde{y}\rangle \tag{2.10}$$

for any closed set $Y \subseteq \mathcal{Y}$, and for any $z \in Z$ and $v \in Y$. In addition, we denote

$$\bar{g}_Y(z) := g_Y(0, z) = \sup_{\tilde{y}\in Y} Q(w^*, x^*, \tilde{y}; z). \tag{2.11}$$

If $Y = \mathcal{Y}$, we will omit the subscript Y and simply use notations $g(v, z)$ and $\bar{g}(z)$.

In Propositions 2.1 and 2.2 below, we describe the relationship between the gap functions (2.10)–(2.11) and the approximate solutions to problems (1.1) and (1.4).

PROPOSITION 2.1. *For any $Y \subseteq \mathcal{Y}$, if $g_Y(Bw - Kx - b, z) \le \varepsilon < \infty$ and $\|Bw - Kx - b\| \le \delta$ where $z = (w, x, y) \in \mathcal{Z}$, then (w, x) is an (ε, δ)-solution of (1.1). In particular, when $Y = \mathcal{Y}$, for any v such that $g(v, z) \le \varepsilon < \infty$ and $\|v\| \le \delta$, we always have $v = Bw - Kx - b$.*

Proof. By (2.8) and (2.10), for all $v \in \mathcal{Y}$ and $Y \subseteq \mathcal{Y}$, we have

$$\begin{aligned} g_Y(v,z) &= \sup_{\tilde{y}\in Y}[G(x)+F(w)-\langle \tilde{y}, Bw-Kx-b\rangle]-[G(x^*)+F(w^*)]+\langle v,\tilde{y}\rangle \\ &= G(x)+F(w)-f^*+\sup_{\tilde{y}\in Y}\langle -\tilde{y}, Bw-Kx-b-v\rangle. \end{aligned}$$

From the above we see that if $g_Y(Bw-Kx-b,z) = G(x)+F(w)-f^* \le \varepsilon$ and $\|Bw-Kx-b\| \le \delta$, then (w,z) is an (ε,δ)-solution. In addition, if $Y=\mathcal{Y}$, we can also see that $g(v,z)=\infty$ if $v \neq Bw-Kx-b$, hence $g(v,z)<\infty$ implies that $v = Bw-Kx-b$. ∎

From Proposition 2.1 we can see that when $Y=\mathcal{Y}$ and $g(v,z)\le\varepsilon$, $\|v\|$ is always the feasibility residual of the approximate solution (w,x). Proposition 2.2 below shows that in some special cases, there exists an approximate solution to problem (1.1) that has zero feasibility residual.

PROPOSITION 2.2. *Assume that B is an one-to-one linear operator such that $B\mathcal{W}=\mathcal{Y}$, and $F(\cdot)$ is Lipschitz continuous, then the set $Y := (B^*)^{-1}\operatorname{dom} F^*$ is bounded. Moreover, if $\bar{g}_Y(z)\le\varepsilon$, then the pair $(\tilde{w},x)$ is an ε-solution of (1.1), where $\tilde{w}=(B^*)^{-1}(Kx+b)$.*

Proof. We can see that $\tilde{w}$ is well-defined since $B\mathcal{W}=\mathcal{Y}$. Also, using the fact that $F(\cdot)$ is finite valued, by Corollary 13.3.3 in [45] we know that $\operatorname{dom} F^*$ *is bounded, hence Y is bounded. In addition, as $B\tilde{w}-Kx-b=0$,*

$$\begin{aligned} \bar{g}_Y(z) &= \sup_{\tilde{y}\in Y}[G(x)+F(w)-\langle \tilde{y}, Bw-Kx-b\rangle]-[G(x^*)+F(w^*)] \\ &= G(x)+F(w)-f^*+\sup_{\tilde{y}\in Y}\langle -\tilde{y}, Bw-B\tilde{w}\rangle \\ &= G(x)+F(\tilde{w})-f^*+\sup_{\tilde{y}\in Y}[F(w)-F(\tilde{w})-\langle B^*\tilde{y}, w-\tilde{w}\rangle]. \end{aligned}$$

If $B^*Y\cap\partial F(\tilde{w})\neq\emptyset$, then from the convexity of $F(\cdot)$ we have

$$\bar{g}_Y(z)\ge G(x)+F(\tilde{w})-f^*,$$

thus $(\tilde{w},x)$ is an ε-solution. To finish the proof it suffices to show that $B^*Y\cap\partial F(\tilde{w})\neq\emptyset$. Observing that

$$\sup_{\bar{w}\in B^*Y}\langle\tilde{w},\bar{w}\rangle-F^*(\bar{w}) = \sup_{\bar{w}\in\operatorname{dom}F^*}\langle\tilde{w},\bar{w}\rangle-F^*(\bar{w}) = \sup_{\bar{w}\in\mathcal{W}}\langle\tilde{w},\bar{w}\rangle-F^*(\bar{w}),$$

and using the fact that Y is closed, we can conclude that there exists $B^*\tilde{y}\in B^*Y$ such that $B^*\tilde{y}$ attains the supremum of the function $\langle\tilde{w},\bar{w}\rangle-F^*(\bar{w})$ with respect to $\bar{w}$. By Theorem 23.5 in [45], we have $B^*\tilde{y}\in\partial F(\tilde{w})$, and hence $\partial F(\tilde{w})\cap B^*Y\neq\emptyset$. ∎

A direct consequence of the above proposition is that for the UCO problem, if $F(\cdot)$ is Lipschitz continuous and $\bar{g}_Y(z)\le\varepsilon$, then (x,Kx) is an ε-solution.

2.2. Main estimations. In this subsection, we present the main estimates that will be used to prove the rate of convergence for AADMM.

LEMMA 2.3. *Let*

$$\Gamma_t = \begin{cases} \Gamma_1 & \text{when } \alpha_t = 1, \\ (1-\alpha_t)\Gamma_{t-1} & \text{when } t>1. \end{cases} \tag{2.12}$$

For all $y \in \mathcal{Y}$, the iterates $\{z_t^{ag}\}_{t\geq 1} := \{(w_t^{ag}, x_t^{ag}, y_t^{ag})\}_{t\geq 1}$ of Algorithm 2 satisfy

$$\begin{aligned}
&\frac{1}{\Gamma_t} Q(w^*, x^*, y; z_{t+1}^{ag}) - \sum_{i=2}^{t} \left(\frac{1-\alpha_i}{\Gamma_i} - \frac{1}{\Gamma_{i-1}} \right) Q(w^*, x^*, y; z_i^{ag}) \\
&\leq \mathcal{B}_t(x^*, x_{[t+1]}, \eta_{[t]}) + \mathcal{B}_t(y, y_{[t+1]}, \rho_{[t]}^{-1}) + \mathcal{B}_t(Bw^*, Bw_{[t+1]}, \theta_{[t]}) - \chi \mathcal{B}_t(Kx^*, Kx_{[t+1]}, \theta_{[t]}) \\
&\quad - \sum_{i=1}^{t} \frac{\alpha_i(\tau_i - \theta_i)}{2\Gamma_i} \|Bw_{i+1} - Kx^* - b\|^2 + \sum_{i=1}^{t} \frac{\alpha_i(\tau_i - \theta_i)}{2\Gamma_i} \|K(x_{i+1} - x^*)\|^2 \\
&\quad - \sum_{i=1}^{t} \frac{\alpha_i(\tau_i - \rho_i)}{2\Gamma_i \rho_i^2} \|y_i - y_{i+1}\|^2 - \sum_{i=1}^{t} \frac{\alpha_i}{2\Gamma_i} \left(\eta_i - L_G \alpha_i - \chi \theta_i \|K\|^2 \right) \|x_i - x_{i+1}\|^2.
\end{aligned} \tag{2.13}$$

where the term $\mathcal{B}_t(\cdot,\cdot,\cdot)$ is defined as follows: for any point v and any sequence $v_{[t+1]}$ in any vectorial space $\mathcal{V}$, and any real valued sequence $\gamma_{[t]}$,

$$\mathcal{B}_t(v, v_{[t+1]}, \gamma_{[t]}) := \sum_{i=1}^{t} \frac{\alpha_i}{2\Gamma_i} \gamma_i \left(\|v_i - v\|^2 - \|v_{i+1} - v\|^2 \right). \tag{2.14}$$

Proof. To start with, we prove an important property of the function $Q(\cdot,\cdot)$ under Algorithm 2. By convexity of $G(\cdot)$ we have

$$G(x_{t+1}^{ag}) \leq \; G(x_t^{md}) + \langle \nabla G(x_t^{md}), x_{t+1}^{ag} - x_t^{md} \rangle + \frac{L_G}{2} \|x_{t+1}^{ag} - x_t^{md}\|^2. \tag{2.15}$$

Moreover, by equations (2.1) and (2.3), $x_{t+1}^{ag} - x_{t+1}^{md} = \alpha_t(x_{t+1} - x_t)$. Using this observation, equation (2.15) and the convexity of $G(\cdot)$, we have

$$\begin{aligned}
&G(x_{t+1}^{ag}) \leq \; G(x_t^{md}) + \langle \nabla G(x_t^{md}), x_{t+1}^{ag} - x_t^{md} \rangle + \frac{L_G \alpha_t^2}{2} \|x_{t+1} - x_t\|^2 \\
&= G(x_t^{md}) + (1-\alpha_t) \langle \nabla G(x_t^{md}), x_t^{ag} - x_t^{md} \rangle + \alpha_t \langle \nabla G(x_t^{md}), x_{t+1} - x_t^{md} \rangle + \frac{L_G \alpha_t^2}{2} \|x_{t+1} - x_t\|^2 \\
&= (1-\alpha_t) \left[G(x_t^{md}) + \langle \nabla G(x_t^{md}), x_t^{ag} - x_t^{md} \rangle \right] + \alpha_t \left[G(x_t^{md}) + \langle \nabla G(x_t^{md}), x_{t+1} - x_t^{md} \rangle \right] \\
&\quad + \frac{L_G \alpha_t^2}{2} \|x_{t+1} - x_t\|^2 \\
&= (1-\alpha_t) \left[G(x_t^{md}) + \langle \nabla G(x_t^{md}), x_t^{ag} - x_t^{md} \rangle \right] + \alpha_t \left[G(x_t^{md}) + \langle \nabla G(x_t^{md}), x - x_t^{md} \rangle \right] \\
&\quad + \alpha_t \langle \nabla G(x_t^{md}), x_{t+1} - x \rangle + \frac{L_G \alpha_t^2}{2} \|x_{t+1} - x_t\|^2 \\
&\leq (1-\alpha_t) G(x_t^{ag}) + \alpha_t G(x) + \alpha_t \langle \nabla G(x_t^{md}), x_{t+1} - x \rangle + \frac{L_G \alpha_t^2}{2} \|x_{t+1} - x_t\|^2, \; \forall x \in X.
\end{aligned} \tag{2.16}$$

By (1.10), (1.11), (2.8), (2.16) and the convexity of $F(\cdot)$, we conclude that

$$\begin{aligned}
&Q(z; z_{t+1}^{ag}) - (1-\alpha_t) Q(z; z_t^{ag}) \\
&= [G(x_{t+1}^{ag}) + F(w_{t+1}^{ag}) - \langle y, Bw_{t+1}^{ag} - Kx_{t+1}^{ag} - b \rangle] - [G(x) + F(w) - \langle y_{t+1}^{ag}, Bw - Kx - b \rangle] \\
&\quad - (1-\alpha_t)[G(x_t^{ag}) + F(w_t^{ag}) - \langle y, Bw_t^{ag} - Kx_t^{ag} - b \rangle] + (1-\alpha_t)[G(x) + F(w) - \langle y_t^{ag}, Bw - Kx - b \rangle] \\
&= \left[G(x_{t+1}^{ag}) - (1-\alpha_t) G(x_t^{ag}) - \alpha_t G(x) \right] + \left[F(w_{t+1}^{ag}) - (1-\alpha_t) F(w_t^{ag}) - \alpha_t F(w) \right] \\
&\quad - \alpha_t \langle y, Bw_{t+1} - Kx_{t+1} - b \rangle + \alpha_t \langle y_{t+1}, Bw - Kx - b \rangle \\
&\leq \alpha_t \Big\{ \langle \nabla G(x_t^{md}), x_{t+1} - x \rangle + [F(w_{t+1}) - F(w)] + \frac{L_G \alpha_t}{2} \|x_{t+1} - x_t\|^2 \\
&\quad - \langle y, Bw_{t+1} - Kx_{t+1} - b \rangle + \langle y_{t+1}, Bw - Kx - b \rangle \Big\}.
\end{aligned} \tag{2.17}$$

Next, we examine the optimality conditions in (2.2) and (2.4). for all $x \in X$ and $w \in \mathcal{W}$, we have

$$\begin{aligned}&\langle \nabla G(x_t^{md}) + \eta_t(x_{t+1} - x_t), x_{t+1} - x\rangle - \langle \theta_t(Bw_t - K\tilde{x}_t - b) - y_t, K(x_{t+1} - x)\rangle \le 0, \text{ and}\\ &F(w_{t+1}) - F(w) + \langle \tau_t(Bw_{t+1} - Kx_{t+1} - b) - y_t, B(w_{t+1} - w)\rangle \le 0,\end{aligned}$$

where

$$\tilde{x}_t := \chi x_t + (1-\chi)x_{t+1}. \tag{2.18}$$

Observing from (2.6) that $Bw_{t+1} - Kx_{t+1} - b = (y_t - y_{t+1})/\rho_t$ and $Bw_t - K\tilde{x}_t - b = (y_t - y_{t+1})/\rho_t - K(\tilde{x}_t - x_{t+1}) + B(w_t - w_{t+1})$, the optimality conditions become

$$\begin{aligned}&\langle \nabla G(x_t^{md}) + \eta_t(x_{t+1} - x_t), x_{t+1} - x\rangle + \langle \left(\frac{\theta_t}{\rho_t} - 1\right)(y_t - y_{t+1}) - y_{t+1}, -K(x_{t+1} - x)\rangle\\ &\quad + \theta_t\langle K(\tilde{x}_t - x_{t+1}), K(x_{t+1} - x)\rangle + \theta_t\langle B(w_t - w_{t+1}), -K(x_{t+1} - x)\rangle \le 0, \text{ and}\\ &F(w_{t+1}) - F(w) + \langle \left(\frac{\tau_t}{\rho_t} - 1\right)(y_t - y_{t+1}) - y_{t+1}, B(w_{t+1} - w)\rangle \le 0.\end{aligned}$$

Therefore,

$$\begin{aligned}&\langle \nabla G(x_t^{md}), x_{t+1} - x\rangle + F(w_{t+1}) - F(w) - \langle y, Bw_{t+1} - Kx_{t+1} - b\rangle + \langle y_{t+1}, Bw - Kx - b\rangle\\ &\le \langle \eta_t(x_t - x_{t+1}), x_{t+1} - x\rangle + \langle y_{t+1} - y, Bw_{t+1} - Kx_{t+1} - b\rangle\\ &\quad - \langle \left(\frac{\theta_t}{\rho_t} - 1\right)(y_t - y_{t+1}), -K(x_{t+1} - x)\rangle - \langle \left(\frac{\tau_t}{\rho_t} - 1\right)(y_t - y_{t+1}), B(w_{t+1} - w)\rangle\\ &\quad + \theta_t\langle K(x_{t+1} - \tilde{x}_t), K(x_{t+1} - x)\rangle + \theta_t\langle B(w_{t+1} - w_t), -K(x_{t+1} - x)\rangle.\end{aligned} \tag{2.19}$$

Three observations on the right hand side of (2.19) are in place. Firstly, by (2.6) we have

$$\begin{aligned}&\langle \eta_t(x_t - x_{t+1}), x_{t+1} - x\rangle + \langle y_{t+1} - y, Bw_{t+1} - Kx_{t+1} - b\rangle\\ &= \eta_t\langle x_t - x_{t+1}, x_{t+1} - x\rangle + \frac{1}{\rho_t}\langle y_{t+1} - y, y_t - y_{t+1}\rangle\\ &= \frac{\eta_t}{2}(\|x_t - x\|^2 - \|x_{t+1} - x\|^2) - \frac{\eta_t}{2}(\|x_t - x_{t+1}\|^2) + \frac{1}{2\rho_t}\left(\|y_t - y\|^2 - \|y_{t+1} - y\|^2 - \|y_t - y_{t+1}\|^2\right),\end{aligned} \tag{2.20}$$

and secondly, by (2.6) we can see that

$$B(w_{t+1} - w) = \frac{1}{\rho_t}(y_t - y_{t+1}) + (Kx_{t+1} - Kx) - (Bw - Kx - b), \tag{2.21}$$

and

$$\begin{aligned}&\langle \left(\frac{\theta_t}{\rho_t} - 1\right)(y_t - y_{t+1}), K(x_{t+1} - x)\rangle - \langle \left(\frac{\tau_t}{\rho_t} - 1\right)(y_t - y_{t+1}), \frac{1}{\rho_t}(y_t - y_{t+1}) + (Kx_{t+1} - Kx)\rangle\\ &= \frac{\tau_t - \theta_t}{\rho_t}\langle y_t - y_{t+1}, -K(x_{t+1} - x)\rangle - \frac{\tau_t - \rho_t}{\rho_t^2}\|y_t - y_{t+1}\|^2\\ &= \frac{\tau_t - \theta_t}{2}\left[\frac{1}{\rho_t^2}\|y_t - y_{t+1}\|^2 + \|K(x_{t+1} - x)\|^2 - \|\frac{1}{\rho_t}(y_t - y_{t+1}) + K(x_{t+1} - x)\|^2\right] - \frac{\tau_t - \rho_t}{\rho_t^2}\|y_t - y_{t+1}\|^2\\ &= \frac{\tau_t - \theta_t}{2}\left[\frac{1}{\rho_t^2}\|y_t - y_{t+1}\|^2 + \|K(x_{t+1} - x)\|^2 - \|Bw_{t+1} - Kx - b\|^2\right] - \frac{\tau_t - \rho_t}{\rho_t^2}\|y_t - y_{t+1}\|^2.\end{aligned} \tag{2.22}$$

Thirdly, from (2.18) *we have*

$$\begin{aligned}
&\theta_t\langle K(x_{t+1}-\tilde{x}_t), K(x_{t+1}-x)\rangle + \theta_t\langle B(w_{t+1}-w_t), -K(x_{t+1}-x)\rangle \\
=\ & -\frac{\chi\theta_t}{2}\left(\|K(x_t-x)\|^2 - \|K(x_{t+1}-x)\|^2 - \|K(x_t-x_{t+1})\|^2\right) \\
& +\frac{\theta_t}{2}\left(\|Bw_t-Kx-b\|^2 - \|Bw_{t+1}-Kx-b\|^2 + \|Bw_{t+1}-Kx_{t+1}-b\|^2 - \|Bw_t-Kx_{t+1}-b\|^2\right) \\
\le\ & -\frac{\chi\theta_t}{2}\left(\|K(x_t-x)\|^2 - \|K(x_{t+1}-x)\|^2\right) + \frac{\chi\theta_t\|K\|^2}{2}\|x_t-x_{t+1}\|^2 \\
& +\frac{\theta_t}{2}\left(\|Bw_t-Kx-b\|^2 - \|Bw_{t+1}-Kx-b\|^2\right) + \frac{\theta_t}{2\rho_t^2}\|y_t-y_{t+1}\|^2 - \frac{\theta_t}{2}\|Bw_t-Kx_{t+1}-b\|^2,
\end{aligned} \tag{2.23}$$

where the last inequality results from the fact that

$$\chi\|K(x_t-x_{t+1})\| \le \chi\|K\|\|x_t-x_{t+1}\|. \tag{2.24}$$

Applying (2.19) – (2.23) *to* (2.17), *we have*

$$\begin{aligned}
&\frac{1}{\Gamma_t}Q(z;z_{t+1}^{ag}) - \frac{1-\alpha_t}{\Gamma_t}Q(z;z_t^{ag}) \\
\le\ & \frac{\alpha_t}{\Gamma_t}\Bigg\{\frac{\eta_t}{2}(\|x_t-x\|^2 - \|x_{t+1}-x\|^2) + \frac{1}{2\rho_t}(\|y_t-y\|^2 - \|y_{t+1}-y\|^2) - \frac{\tau_t-\rho_t}{2\rho_t^2}\|y_t-y_{t+1}\|^2 \\
& +\frac{\theta_t}{2}\|Bw_t-Kx-b\|^2 - \frac{\tau_t}{2}\|Bw_{t+1}-Kx-b\|^2 - \frac{\chi\theta_t}{2}(\|K(x_t-x)\|^2 - \|K(x_{t+1}-x)\|^2) \\
& +\langle\left(\frac{\tau_t}{\rho_t}-1\right)(y_t-y_{t+1}), Bw-Kx-b\rangle + \frac{\tau_t-\theta_t}{2}\|K(x_{t+1}-x)\|^2 - \frac{\theta_t}{2}\|Bw_t-Kx_{t+1}-b\|^2 \\
& -\frac{1}{2}\left(\eta_t - L_G\alpha_t - \chi\theta_t\|K\|^2\right)\|x_t-x_{t+1}\|^2\Bigg\}.
\end{aligned} \tag{2.25}$$

Letting $w=w^*$ *and* $x=x^*$ *in the above, observing from* (2.12) *that* $\Gamma_{t-1}=(1-\alpha_t)/\Gamma_t$, *in view of* (2.14) *and applying the above inequality inductively, we conclude* (2.13). □

There are two major consequences of Lemma 2.3. If $\alpha_t \equiv 1$ for all t, then the left hand side of (2.13) becomes $\frac{1}{\Gamma_1}\sum_{i=2}^t Q(z;z_i^{ag})$. On the other hand, if $\alpha_t \in [0,1)$ for all t, then in view of (2.12), the left hand side of (2.13) is $Q(z;z_{t+1}^{ag})/\Gamma_t$. This difference is the main reason why we can accelerate the rate of convergence of AADMM in terms of L_G.

In the next lemma, we provide possible bounds of $\mathcal{B}(\cdot,\cdot,\cdot)$ in Lemma 2.3.

LEMMA 2.4. *Suppose that* $\mathcal{V}$ *is any vector space and* $V\subset\mathcal{V}$ *is any convex set. For any* $v\in V$, $v_{[t+1]}\subset\mathcal{V}$ *and* $\gamma_{[t]}\subset\mathbb{R}$, *we have the following:*

a). If the sequence $\{\alpha_i\gamma_i/\Gamma_i\}$ *is decreasing, then*

$$\mathcal{B}_t(v,v_{[t+1]},\gamma_{[t]}) \le \frac{\alpha_1\gamma_1}{2\Gamma_1}\|v_1-v\|^2 - \frac{\alpha_t\gamma_t}{2\Gamma_t}\|v_{t+1}-v\|^2. \tag{2.26}$$

b). If the sequence $\{\alpha_i\gamma_i/\Gamma_i\}$ *is increasing,* V *is bounded and* $v_{[t+1]}\subset V$, *then*

$$\mathcal{B}_t(v,v_{[t+1]},\gamma_{[t]}) \le \frac{\alpha_t\gamma_t}{2\Gamma_t}D_V^2 - \frac{\alpha_t\gamma_t}{2\Gamma_t}\|v_{t+1}-v\|^2. \tag{2.27}$$

Proof. By (2.14) *we have*

$$\mathcal{B}_t(v,v_{[t+1]},\gamma_{[t]}) = \frac{\alpha_1\gamma_1}{2\Gamma_1}\|v_1-v\|^2 - \sum_{i=1}^{t-1}\left(\frac{\alpha_i\gamma_i}{2\Gamma_i} - \frac{\alpha_{i+1}\gamma_{i+1}}{2\Gamma_{i+1}}\right)\|v_{i+1}-v\|^2 - \frac{\alpha_t\gamma_t}{2\Gamma_t}\|v_{t+1}-v\|^2.$$

If the sequence $\{\alpha_i\gamma_i/\Gamma_i\}$ is decreasing, then the above equation implies (2.26). If the sequence $\{\alpha_i\gamma_i/\Gamma_i\}$ is increasing, V is bounded and $v_{[t+1]} \subset V$, then from the above equation we have

$$\begin{aligned}\mathcal{B}_t(v, v_{[t+1]}, \gamma_{[t]}) &\le \frac{\alpha_1\gamma_1}{2\Gamma_1}D_V^2 - \sum_{i=1}^{t-1}\left(\frac{\alpha_i\gamma_i}{2\Gamma_i} - \frac{\alpha_{i+1}\gamma_{i+1}}{2\Gamma_{i+1}}\right)D_V^2 - \frac{\alpha_t\gamma_t}{2\Gamma_t}\|v_{t+1} - v\|^2 \\ &= \frac{\alpha_t\gamma_t}{2\Gamma_t}D_V^2 - \frac{\alpha_t\gamma_t}{2\Gamma_t}\|v_{t+1} - v\|^2,\end{aligned}$$

hence (2.27) holds. □

2.3. Convergence results on solving UCO problems in bounded domain. We study UCO problems with bounded feasible sets in this subsection. In particular, throughout this subsection we assume that

$$\textit{Both } X \textit{ and } Y := \operatorname{dom} F^* \textit{ are compact, and } B = I,\ b = 0. \tag{2.28}$$

It should be noted that the boundedness of Y above is equivalent to the Lipschitz continuity of $F(\cdot)$ (see, e.g, Corollary 13.3.3 in [45]).

The following Theorem 2.5 generalizes the convergence properties of ADMM algorithms. Although the convergence analysis of ADMM, L-ADMM and P-ADMM has already been done in several literatures (e.g., [34, 25, 10, 42]), Theorem 2.5 gives a unified view of the convergence properties of all ADMM algorithms.

THEOREM 2.5. *In AADMM, if the parameters of are set to $\alpha_t \equiv 1$, $\theta_t \equiv \tau_t \equiv \rho_t \equiv \rho$ and $\eta_t \equiv L_G + \chi\rho\|K\|^2$, then*

$$G(x^{t+1}) + F(Kx^{t+1}) - f^* \le \frac{L_G}{2t}D_X^2 + \frac{\chi\rho}{2t}\|K\|^2 D_X^2 + \frac{(1-\chi)\rho}{2t}D_{X,K}^2 + \frac{D_Y^2}{2\rho t}, \tag{2.29}$$

where $x^{t+1} := \dfrac{1}{t}\displaystyle\sum_{i=2}^{t+1} x_i$. Specially, if ρ is given by

$$\rho = \frac{D_Y}{\chi\|K\|D_X + (1-\chi)D_{X,K}}, \tag{2.30}$$

then

$$G(x^{t+1}) + F(\tilde{w}^{t+1}) - f^* \le \frac{L_G D_X^2}{2t} + \frac{\chi\|K\|D_X D_Y + (1-\chi)D_{X,K}D_Y}{t}. \tag{2.31}$$

Proof. Since $\alpha_t \equiv 1$, By (2.3), (2.5) and (2.7) we have $x_t^{ag} = x_t$, $w_t^{ag} = w_t$ and $y_t^{ag} = y_t$, and we can see that $\Gamma_t \equiv 1$ satisfies (2.12) . Applying the parameter settings to RHS of (2.13) in Lemma 2.3, we have

$$\begin{aligned}\mathcal{B}_t(x^*, x_{[t+1]}, \eta_{[t]}) &= \frac{\eta}{2}(\|x_1 - x^*\|^2 - \|x_{t+1} - x^*\|^2) \\ &\le \frac{L_G}{2}D_{x^*}^2 + \frac{\chi\rho}{2}\|K\|^2 D_{x^*}^2 - \frac{\chi\rho}{2}\|K\|^2\|x_{t+1} - x^*\|^2, \\ \mathcal{B}_t(w^*, w_{[t+1]}, \theta_{[t]}) &= \frac{\rho}{2}(\|w_1 - w^*\|^2 - \|w_{t+1} - w^*\|^2) \le \frac{\rho D_{w^*}^2}{2} = \frac{\rho D_{x^*,K}^2}{2}, \\ -\chi\mathcal{B}_t(Kx^*, Kx_{[t+1]}, \theta_{[t]}) &= -\frac{\chi\rho}{2}(\|Kx_1 - Kx^*\|^2 - \|Kx_{t+1} - Kx^*\|^2) \\ &\le -\frac{\chi\rho}{2}D_{x^*,K}^2 + \frac{\chi\rho}{2}\|K\|^2\|x_{t+1} - x^*\|^2, \\ \mathcal{B}_t(y, y_{[t+1]}, \rho_{[t]}^{-1}) &\le \frac{1}{2\rho}(\|y_1 - y\|^2 - \|y_{t+1} - y\|^2) \le \frac{D_Y^2}{2\rho},\ \forall y \in Y.\end{aligned}$$

Therefore, by Lemma 2.3 we have

$$\sum_{i=2}^{t+1} Q(w^*, x^*, y; z_i) \le \frac{L_G}{2} D_{x^*}^2 + \frac{\chi\rho}{2}\|K\|^2 D_{x^*}^2 + \frac{(1-\chi)\rho}{2} D_{x^*,K}^2 + \frac{D_Y^2}{2\rho}$$
$$\le \frac{L_G}{2} D_X^2 + \frac{\chi\rho}{2}\|K\|^2 D_X^2 + \frac{(1-\chi)\rho}{2} D_{X,K}^2 + \frac{D_Y^2}{2\rho},\ \forall y \in Y.$$

Furthermore, noticing that for all $y \in Y$, by the convexity of $Q(x^*, w^*, y; \cdot)$,

$$Q(w^*, x^*, y; z^{t+1}) \le \frac{1}{t}\sum_{i=2}^{t+1} Q(w^*, x^*, y; z_i),\ \text{where } z^{t+1} := \frac{1}{t}\sum_{i=2}^{t+1} z_i.$$

Applying the two inequalities above to (2.11) and Proposition 2.2, we conclude (2.29), and (2.31) follows immediately. □

Although AADMM unifies all ADMM algorithms, what makes it most special is the variable weighting sequence $\{\alpha_t\}_{t\ge1}$ (rather than $\alpha_t = 1$) that accelerates its convergence rate with respect to its dependence on L_G, as shown in Theorem 2.6 below.

THEOREM 2.6. *In AADMM, if the parameters are set to*

$$\alpha_t = \frac{2}{t+1}, \tau_t = \rho_t \equiv \rho,\ \theta_t = \frac{(t-1)\rho}{t},\ \text{and } \eta_t = \frac{2L_G + \chi\rho t\|K\|^2}{t}, \tag{2.32}$$

then

$$G(x_{t+1}^{ag}) + F(Kx_{t+1}^{ag}) - f^* \le \frac{2L_G D_X^2}{t(t+1)} + \frac{1}{t+1}\left[\chi\rho\|K\|^2 D_X^2 + (1-\chi)\rho D_{X,K}^2 + \frac{D_Y^2}{\rho}\right]. \tag{2.33}$$

In particular, if ρ is given by (2.30), then

$$G(x_{t+1}^{ag}) + F(Kx_{t+1}^{ag}) - f^* \le \frac{2L_G D_X^2}{t(t+1)} + \frac{2}{t+1}\left[\chi\|K\| D_X D_Y + (1-\chi) D_{X,K} D_Y\right]. \tag{2.34}$$

Proof. It is clear that

$$\alpha_t = \frac{2}{t+1}\ \text{and}\ \Gamma_t = \frac{2}{t(t+1)}\ \text{satisfies (2.12), and}\ \frac{\alpha_t}{\Gamma_t} = t. \tag{2.35}$$

By the parameter setting (2.32) and the definition of $\mathcal{B}(\cdot,\cdot,\cdot)$ in (2.14), it is easy to calculate that

$$\eta_t - L_G\alpha_t - \chi\theta_t\|K\|^2 \ge 0,\ \tau_t \ge \theta_t,$$
$$\mathcal{B}_t(w^*, w_{[t+1]}, \theta_t) - \sum_{i=1}^{t}\frac{\alpha_i(\tau_i - \theta_i)}{2\Gamma_i}\|w_{i+1} - w^*\|^2 = -\frac{\rho t}{2}\|w_{t+1} - w^*\|^2 \le 0,$$
$$-\chi\mathcal{B}_t(Kx^*, Kx_{[t+1]}, \theta_t) + \sum_{i=1}^{t}\frac{\alpha_i(\tau_i - \theta_i)}{2\Gamma_i}\|Kx_{i+1} - Kx^*\|^2$$
$$= \frac{\chi\rho t}{2}\|Kx_{t+1} - Kx^*\|^2 + \frac{(1-\chi)\rho}{2}\sum_{i=1}^{t}\|Kx_{i+1} - Kx^*\|^2 \le \frac{\chi\rho t}{2}\|K\|^2\|x_{t+1} - x^*\|^2 + \frac{(1-\chi)\rho t}{2} D_{X,K}^2.$$

Moreover, by (2.4), (2.6) and Moreau's decomposition theorem (see, e.g., [37, 14, 18]), we have

$$\begin{aligned} y_{t+1} &= y_t - \rho(w_{t+1} - Kx_{t+1} - b) \\ &= (y_t + \rho Kx_{t+1} + \rho b) - \rho\operatorname*{argmin}_{w\in\mathcal{W}} F(w) + \frac{\rho}{2}\|w - \frac{1}{\rho}(y_t - Kx_{t+1} - b)\|^2 \\ &= \operatorname*{argmin}_{y\in\mathcal{Y}} F^*(y) + \frac{1}{2\rho}\|y - \frac{1}{\rho}(y_t - Kx_{t+1} - b)\|^2, \end{aligned} \tag{2.36}$$

which implies that $y_{[t+1]} \subset Y$. *Using this observation together with the fact that* $\alpha_t/(\Gamma_t\rho_t) = t/\rho$, *and applying* (2.27) *in Lemma 2.4, we obtain*

$$\mathcal{B}_t(y, y_{[t+1]}, \rho_{[t]}^{-1}) \le \frac{t}{2\rho} D_Y^2, \ \forall y \in Y.$$

Finally, noting that $\alpha_t\eta_t/\Gamma_t = 2L_G + \chi\rho t\|K\|^2$, *by* (2.27) *in Lemma 2.4 we have*

$$\mathcal{B}_t(x^*, x_{[t+1]}, \eta_{[t]}) \le \frac{\alpha_t\eta_t}{2\Gamma_t} D_X^2 - \frac{\alpha_t\eta_t}{2\Gamma_t}\|x_{t+1} - x^*\|^2 \le L_G D_X^2 + \frac{\chi\rho t}{2}\|K\|^2 D_X^2 - \frac{\chi\rho t}{2}\|K\|^2\|x_{t+1} - x^*\|^2.$$

Applying all above inequalities to (2.13) *in Lemma 2.3, we have*

$$\frac{1}{\Gamma_t} Q(w^*, x^*, y; z_{t+1}^{ag}) \le L_G D_X^2 + \frac{\chi\rho t}{2}\|K\|^2 D_X^2 + \frac{(1-\chi)\rho t}{2} D_{X,K}^2 + \frac{t}{2\rho} D_Y^2, \ \forall y \in Y.$$

Using (2.35) *and applying Proposition 2.2, we conclude* (2.33), *and* (2.34) *comes from* (2.30) *and* (2.33). ∎

In view of Theorems 2.5 and 2.6, several remarks on the AADMM algorithms are in place. Firstly, Theorem 2.6 provides an example of choosing stepsizes in AL-ADMM and ALP-ADMM, that leads to better convergence properties w.r.t the dependence on L_G than L-ADMM and LP-ADMM respectively. In particular, AL-ADMM and ALP-ADMM allow L_G to be as large as $\Omega(N)$ without affecting the rate of convergence (up to a constant factor). The comparison of these AADMM algorithms in terms of their rates of convergence is shown in Table 2.1. Secondly, ALP-ADMM has the same rate of convergence as Nesterov's smoothing scheme [40], and achieves optimal rate of convergence (1.16). Moreover, we can see from (2.36) that the APD method in [13] is equivalent to ALP-ADMM. Nonetheless, AL-ADMM has better constant in the estimation of rate of convergence than both ALP-ADMM and Nesterov's smoothing scheme, since $D_{X,K} \le \|K\|D_X$. However, the computational time for solving problem (2.2) with $\chi = 0$ is usually higher than that for $\chi = 1$, hence AL-ADMM has higher iteration cost than that of ALP-ADMM. The trade-off between better rate constants and cheaper iteration costs has to be considered in practice. Thirdly, while Theorem 2.5 describes only the ergodic convergence of the ADMM algorithms, Theorem 2.6 describes the convergence of aggregate sequences $\{z_{t+1}^{ag}\}_{t\ge1}$, which are exactly the outputs of the accelerated schemes. Finally, in ADMM methods we have $\tau_t = \rho_t = \theta_t$, while in Theorem 2.6 we only have $\tau_t = \rho_t$, although $\theta_t \to \rho_t$ when $t \to \infty$. In fact, if the total number of iterations is given, it is possible to choose a set of equal stepsize parameters, as described by Theorem 2.7 below.

THEOREM 2.7. *In AADMM, if the total number of iterations* N *is chosen, and the parameters are set to*

$$\alpha_t = \frac{2}{t+1}, \ \theta_t = \tau_t = \rho_t = \frac{\rho N}{t}, \ \text{and } \eta_t = \frac{2L_G + \chi\rho N\|K\|^2}{t},$$

where ρ *is given by* (2.30), *then*

$$G(x_N^{ag}) + F(Kx_N^{ag}) - f^* \le \frac{2L_G D_X^2}{N(N-1)} + \frac{2}{N-1}\left[\chi\|K\|D_X D_Y + (1-\chi)D_{X,K} D_Y\right]. \tag{2.37}$$

Proof. *Using equation* (2.35) *as well as the definition of* $\mathcal{B}(\cdot,\cdot,\cdot)$ *in* (2.14)*, it is easy to calculate that*

$$\begin{aligned}
\eta_t - L_G\alpha_t - \chi\theta_t\|K\|^2 &\geq 0,\\
\mathcal{B}_t(x^*, x_{[t+1]}, \eta_{[t]}) &= \frac{2L_G + \chi\rho N\|K\|^2}{2}(\|x_1 - x^*\|^2 - \|x_{t+1} - x^*\|^2)\\
&\leq \frac{2L_G + \chi\rho N\|K\|^2}{2}(D_{x^*}^2 - \|x_{t+1} - x^*\|^2),\\
\mathcal{B}_t(w^*, w_{[t+1]}, \theta_{[t]}) &= \frac{\rho N}{2}(\|w_1 - w^*\|^2 - \|w_{t+1} - w^*\|^2) \leq \frac{\rho N D_{w^*}^2}{2} = \frac{\rho N D_{x^*,K}^2}{2}, \text{ and}\\
-\chi\mathcal{B}_t(Kx^*, Kx_{[t+1]}, \theta_{[t]}) &= -\frac{\chi\rho N}{2}(\|Kx_1 - Kx^*\|^2 - \|Kx_{t+1} - Kx^*\|^2)\\
&\leq -\frac{\chi\rho N}{2}D_{x^*,K}^2 + \frac{\chi\rho N\|K\|^2}{2}\|x_{t+1} - x^*\|^2.
\end{aligned}$$

On the other hand, noting that $\alpha_t/(\Gamma_t\rho_t) = t^2/(\rho N)$*, by* (2.27) *in Lemma* 2.4 *we have*

$$\mathcal{B}_t(y, y_{[t+1]}, \rho_{[t]}^{-1}) \leq \frac{t^2}{2\rho N}(D_Y^2 - \|y_{t+1} - y^*\|^2) \leq \frac{t^2 D_Y^2}{2\rho N} \leq \frac{N}{2\rho}D_Y^2, \ \forall y \in Y, \forall t \leq N.$$

Applying all the above inequalities to (2.13) *in Lemma* 2.3*, we conclude*

$$\begin{aligned}
&\frac{1}{\Gamma_t}Q(w^*, x^*, y; z_{t+1}^{ag}) \leq L_G D_{x^*}^2 + \frac{\chi\rho N}{2}\|K\|^2 D_{x^*}^2 + \frac{(1-\chi)\rho N}{2}D_{x^*,K}^2 + \frac{N}{2\rho}D_Y^2,\\
&\leq L_G D_X^2 + \frac{\chi\rho N}{2}\|K\|^2 D_X^2 + \frac{(1-\chi)\rho N}{2}D_{X,K}^2 + \frac{N}{2\rho}D_Y^2.
\end{aligned}$$

Setting $t = N - 1$*, and applying* (2.35), (2.30) *and the above inequality to Proposition* 2.2*, we obtain* (2.37). ∎

Table 2.1: Rates of convergence of instances of AADMM for solving UCO with bounded feasible set

	No preconditioning ($\chi = 0$)	Preconditioned ($\chi = 1$)
ADMM	$\mathcal{O}\left(\frac{D_{X,K}D_Y}{t}\right)$	$\mathcal{O}\left(\frac{\|K\|D_X D_Y}{t}\right)$
Linearized ADMM	$\mathcal{O}\left(\frac{L_G D_X^2}{t} + \frac{D_{X,K}D_Y}{t}\right)$	$\mathcal{O}\left(\frac{L_G D_X^2}{t} + \frac{\|K\|D_X D_Y}{t}\right)$
Accelerated	$\mathcal{O}\left(\frac{L_G D_X^2}{t^2} + \frac{D_{X,K}D_Y}{t}\right)$	$\mathcal{O}\left(\frac{L_G D_X^2}{t^2} + \frac{\|K\|D_X D_Y}{t}\right)$

2.4. Convergence results on solving AECCO problems. In this section, we study the rate of convergence of AADMM for solving general AECCO problems without boundedness assumption for either X or Y, in terms of both primal and feasibility residuals. We start with the convergence analysis of ADMM algorithms as a special case of AADMM where $\alpha_t = 1$, $\theta_t = \tau_t = \rho_t = \rho$.

THEOREM 2.8. *In AADMM, if* $\alpha_t \equiv 1$, $\theta_t \equiv \tau_t \equiv \rho_t \equiv \rho$ *and* $\eta_t \equiv \eta \geq L_G + \chi\rho\|K\|^2$*, then*

$$G(x^{t+1}) + F(w^{t+1}) - f^* \leq \frac{1}{2t}\left(\eta D_{x^*}^2 + \rho(1-\chi)D_{x^*,K}^2\right) \tag{2.38}$$

and

$$\|Bw^{t+1} - Kx^{t+1} - b\|^2 \leq \frac{2}{t^2}\left(\frac{2D_{y^*}^2}{\rho^2} + \frac{\eta D_{x^*}^2}{\rho} + (1-\chi)D_{x^*,K}^2\right), \tag{2.39}$$

where $x^{t+1} := \frac{1}{t}\sum_{i=2}^{t+1} x_i$ *and* $w^{t+1} := \frac{1}{t}\sum_{i=2}^{t+1} w_i$. *Specially, if* $\rho = 1$ *and* $\eta = L_G + \chi\|K\|^2$, *then*

$$G(x^{t+1}) + F(w^{t+1}) - f^* \leq \frac{1}{2t}(L_G D_{x^*}^2 + \chi\|K\|^2 D_{x^*}^2 + (1-\chi)D_{x^*,K}^2) \tag{2.40}$$

and

$$\|Bw^{t+1} - Kx^{t+1} - b\| \leq \frac{2\sqrt{L_G}D_{x^*}}{t} + \frac{\chi\sqrt{2}\|K\|D_{x^*}}{t} + \frac{(1-\chi)\sqrt{2}D_{x^*,K}}{t} + \frac{2D_{y^*}}{t}. \tag{2.41}$$

Proof. Similar as the proof of Theorem 2.5, we have

$$\begin{aligned}
&Q(w^*, x^*, y; z^{t+1}) \\
&\leq \frac{1}{2t}\left[L_G D_{x^*}^2 + \chi\rho\|K\|^2 D_{x^*}^2 + (1-\chi)\rho D_{x^*,K}^2 + \frac{1}{\rho}(\|y_1 - y\|^2 - \|y_{t+1} - y\|^2)\right] &(2.42)\\
&\leq \frac{1}{2t}\left[L_G D_{x^*}^2 + \chi\rho\|K\|^2 D_{x^*}^2 + (1-\chi)\rho D_{x^*,K}^2\right] - \langle\frac{1}{\rho t}(y_1 - y_{t+1}), y\rangle, &(2.43)
\end{aligned}$$

where $z^{t+1} = \sum_{t=2}^{t+1} z_i$. *Noting that* $Q(z^*, z^{t+1}) \geq 0$, *by (2.42) we have*

$$\|y_{t+1} - y^*\|^2 \leq \rho L_G D_{x^*}^2 + \chi\rho^2\|K\|^2 D_{x^*}^2 + (1-\chi)\rho^2 D_{x^*,K}^2 + D_{y^*}^2,$$

hence if we let $v_{t+1} = (y_1 - y_{t+1})/(\rho t)$, *then we have*

$$\begin{aligned}
\|v_{t+1}\|^2 &\leq \frac{2}{\rho^2 t^2}(\|y_1 - y^*\|^2 + \|y_{t+1} - y^*\|^2) \\
&\leq \frac{2}{t^2}(\frac{L_G D_{x^*}^2}{\rho} + \chi\|K\|^2 D_{x^*}^2 + (1-\chi)D_{x^*,K}^2 + \frac{2}{\rho^2}D_{y^*}^2).
\end{aligned}$$

Furthermore, by (2.43) we have

$$g(v_{t+1}, z^{t+1}) \leq \frac{1}{2t}\left[L_G D_{x^*}^2 + \chi\rho\|K\|^2 D_{x^*}^2 + (1-\chi)\rho D_{x^*,K}^2\right].$$

Applying the two inequalities above to Proposition 2.1 we obtain (2.38) and (2.39). The results in (2.45) and (2.46) then follows immediately. □

From Theorem 2.8 we see that the for ADMM algorithms, the rate of convergence of both primal and feasibility residuals are of order $\mathcal{O}(1/t)$. The detailed rate of convergence of each algorithm is listed in Tables 2.2 and 2.3. We observe that a larger value of ρ will increase the right side of (2.38), but decrease that of (2.39). Hence, an "optimal" selection of ρ will be determined by considering both primal and feasibility residuals together. For the sake of simplicity, we set $\rho = 1$.

In Theorem 2.9 below, we show that there exists a weighting sequence $\{\alpha_t\}_{t\geq 1}$ that improves the rate of convergence of Algorithm 2 in terms of its dependence on L_G.

THEOREM 2.9. *In AADMM, if the total number of iterations is set to* N, *and the parameters are set to*

$$\alpha_t = \frac{2}{t+1},\ \theta_t = \tau_t = \frac{N}{t},\ \rho_t = \frac{t}{N},\ \text{and } \eta_t = \frac{2L_G + \chi N\|K\|^2}{t}, \tag{2.44}$$

then

$$G(x_N^{ag}) + F(w_N^{ag}) - f^* \le \frac{2L_G D_{x^*}^2}{N(N-1)} + \frac{1}{2(N-1)}\left[\chi\|K\|^2 D_{x^*}^2 + (1-\chi)D_{x^*,K}^2\right], \tag{2.45}$$

and

$$\|Bw_N^{ag} - Kx_N^{ag} - b\| \le \frac{4\sqrt{L_G}D_{x^*}}{(N-1)\sqrt{N}} + \frac{2\sqrt{2}\chi\|K\|D_{x^*}}{N-1} + \frac{2\sqrt{2}(1-\chi)D_{x^*,K}}{N-1} + \frac{4D_{y^*}}{N-1}. \tag{2.46}$$

Proof. Using equations (2.44), (2.35) *and* (2.14), *we can calculate that*

$$\begin{aligned}
&\eta_t - L_G\alpha_t - \chi\theta_t\|K\|^2 \ge 0,\ \tau_t \ge \rho_t \text{ for all } t \le N,\\
\mathcal{B}_t(x^*, x_{[t+1]}, \eta_{[t]}) &= \frac{2L_G + \chi N\|K\|^2}{2}(D_{x^*}^2 - \|x_{t+1} - x^*\|^2),\\
\mathcal{B}_t(y, y_{[t+1]}, \rho_{[t]}^{-1}) &= \frac{N}{2}(\|y_1 - y\|^2 - \|y_{t+1} - y\|^2),\ \forall y \in \mathcal{Y},\\
\mathcal{B}_t(Bw^*, Bw_{[t+1]}, \theta_{[t]}) &= \frac{N}{2}(\|Bw_1 - Bw^*\|^2 - \|B_{t+1} - Bw^*\|^2) \le \frac{N}{2}D_{w^*,B}^2 = \frac{N}{2}D_{x^*,K}^2,\\
-\chi\mathcal{B}_t(Kx^*, Kx_{[t+1]}, \theta_{[t]}) &= -\frac{\chi N}{2}(\|Kx_1 - Kx^*\|^2 - \|Kx_{t+1} - Kx^*\|^2),\\
&\le -\frac{\chi N}{2}(D_{x^*,K} - \|K\|^2\|x_{t+1} - x^*\|^2).
\end{aligned}$$

Applying all the above calculations to (2.13) *in Lemma 2.3, we have*

$$\begin{aligned}
&\frac{1}{\Gamma_t}Q(w^*, x^*, y; z_{t+1}^{ag})\\
&\le L_G D_{x^*}^2 + \frac{\chi N}{2}\|K\|^2 D_{x^*}^2 + \frac{(1-\chi)N}{2}D_{x^*,K}^2 + \frac{N}{2}(\|y_1 - y\|^2 - \|y_{t+1} - y\|^2),\ \forall y \in \mathcal{Y}.
\end{aligned}$$

Two consequences to the above estimation can be derived. Firstly, since $Q(z^*; z_{t+1}^{ag}) \ge 0$, *we have*

$$\|y_{t+1} - y^*\|^2 \le \frac{2L_G}{N}D_{x^*}^2 + \chi\|K\|^2 D_{x^*}^2 + (1-\chi)D_{x^*,K}^2 + D_{y^*}^2,$$

and

$$\|y_1 - y_{t+1}\|^2 \le 2(\|y_1 - y^*\|^2 + \|y_{t+1} - y^*\|^2) \le \frac{4L_G}{N}D_{x^*}^2 + 2\chi\|K\|^2 D_{x^*}^2 + 2(1-\chi)D_{x^*,K}^2 + 4D_{y^*}^2.$$

Secondly, since $\|y_1 - y\|^2 - \|y_{t+1} - y\|^2 = \|y_1\|^2 - \|y_{t+1}\|^2 - 2\langle y_1 - y_{t+1}, y\rangle \le -2\langle y_1 - y_{t+1}, y\rangle$,

$$\frac{1}{\Gamma_t}Q(w^*, x^*, y; z_{t+1}^{ag}) + N\langle y_1 - y_{t+1}, y\rangle \le L_G D_{x^*}^2 + \frac{\chi N}{2}\|K\|^2 D_{x^*}^2 + \frac{(1-\chi)N}{2}D_{x^*,K}^2,\ \forall y \in \mathcal{Y}.$$

Letting $t = N-1$ *and* $v_N := 2(y_1 - y_{t+1})/(N-1)$, *and applying* (2.35) *and the two above inequalities to Proposition 2.1, we obtain* (2.45) *and* (2.46). ∎

Comparing (2.40) and (2.41) with (2.45) and (2.46) respectively, AL-ADMM and ALP-ADMM are better than both L-ADMM and LP-ADMM respectively, in terms of their rates of convergence of both primal and feasibility residuals. The rates of convergence of AADMM algorithms are outlined in Tables 2.2 and 2.3.

Table 2.2: Rates of convergence of the primal residuals of AADMM instances for solving general AECCO

	No preconditioning ($\chi = 0$)	Preconditioned ($\chi = 1$)
ADMM	$\mathcal{O}\left(\frac{D_{x^*,K}^2}{N}\right)$	$\mathcal{O}\left(\frac{\|K\|D_{x^*}^2}{N}\right)$
Linearized ADMM	$\mathcal{O}\left(\frac{L_G D_{x^*}^2 + D_{x^*,K}^2}{N}\right)$	$\mathcal{O}\left(\frac{L_G D_{x^*}^2 + \|K\|D_{x^*}^2}{N}\right)$
Accelerated	$\mathcal{O}\left(\frac{L_G D_{x^*}^2}{N^2} + \frac{D_{x^*,K}^2}{N}\right)$	$\mathcal{O}\left(\frac{L_G D_{x^*}^2}{N^2} + \frac{\|K\|D_{x^*}^2}{N}\right)$

Table 2.3: Rates of convergence of the feasibility residuals of AADMM instances for solving general AECCO

	No preconditioning ($\chi = 0$)	Preconditioned ($\chi = 1$)
ADMM	$\mathcal{O}\left(\frac{D_{x^*,K} + D_{y^*}}{N}\right)$	$\mathcal{O}\left(\frac{\|K\|D_{x^*} + D_{y^*}}{N}\right)$
Linearized ADMM	$\mathcal{O}\left(\frac{\sqrt{L_G}D_{x^*} + D_{x^*,K} + D_{y^*}}{N}\right)$	$\mathcal{O}\left(\frac{\sqrt{L_G}D_{x^*} + \|K\|D_{x^*} + D_{y^*}}{N}\right)$
Accelerated	$\mathcal{O}\left(\frac{\sqrt{L_G}D_{x^*}}{N^{3/2}} + \frac{D_{x^*,K} + D_{y^*}}{N}\right)$	$\mathcal{O}\left(\frac{\sqrt{L_G}D_{x^*}}{N^{3/2}} + \frac{\|K\|D_{x^*} + D_{y^*}}{N}\right)$

2.5. A simple backtracking scheme. We have discussed the rate of convergence of Algorithm 2, with the assumption that both L_G and $\|K\|$ are given. In practice, we may need backtracking techniques to estimate both constants. In this subsection, we propose a simple backtracking technique for AL-ADMM and ALP-ADMM.

From the proof of Lemma 2.3, we can see that if L_G and $\|K\|$ in (2.15) and (2.24) are replaced by L_t and M_t respectively, i.e.,

$$G(x_{t+1}^{ag}) \le G(x_t^{md}) + \langle \nabla G(x_t^{md}), x_{t+1}^{ag} - x_t^{md} \rangle + \frac{L_t}{2}\|x_{t+1}^{ag} - x_t^{md}\|^2 \text{ and} \tag{2.47}$$

$$\chi\|K(x_t - x_{t+1})\| \le \chi M_t \|x_t - x_{t+1}\|, \tag{2.48}$$

then Lemma 2.3 still holds. On the other hand, to prove Theorems 2.5 through 2.9, in addition to Lemma 2.3, we require monotonicity of the sequences $\alpha_{[t]}\eta_{[t]}/\Gamma_{[t]}$, $\alpha_{[t]}\tau_{[t]}/\Gamma_{[t]}$, $\alpha_{[t]}\theta_{[t]}/\Gamma_t$ and $\alpha_t/(\Gamma_t\rho_t)$, and

$$\eta_t - L_t\alpha_t - \chi\theta_t M_t^2 \ge 0, \tag{2.49}$$

The monotonicity of these sequences is also used in Lemma 2.4, which helps to prove the boundedness of distances $\mathcal{B}(\cdot,\cdot,\cdot)$ at the RHS of (2.13) in Lemma 2.3. From these observations, we can simply use the following choice of parameters:

$$\theta_t = \tau_t = \frac{\nu_t\alpha_t}{\Gamma_t},\ \rho_t = \frac{\alpha_t}{\nu_t\Gamma_t},\ \eta_t = L_t\alpha_t + \chi\theta_t M_t^2,$$

where we assume that $\nu_{[t]}$, $M_{[t]}$ are both monotone. It should be noted that the monotonicity of $\alpha_{[t]}\eta_{[t]}/\Gamma_{[t]}$ relies on $\{L_t\alpha_t^2/\Gamma_t\}_{t\ge1}$, which is trivial if we simply set $L_t\alpha_t^2 = \Gamma_t$. In addition, in view of the RHS of (2.13),

we require $\tau_t \geq \rho_t$, i.e., $\nu_t \geq \alpha_t/\Gamma_t$. We summarize all the discussions above to a simple backtracking procedure below.

Procedure 1 Backtracking procedure for AL-ADMM and ALP-ADMM at the t-th iteration

0: **procedure** BACKTRACKING(L_{t-1}, M_{t-1}, Γ_{t-1}, ν_{t-1}, x_t, x_t^{ag}, L_{min})
1: $L_t \leftarrow \max\{L_{min}, L_{t-1}/2\}$, $M_t = M_{t-1}$ and $v_t = v_{t-1}$. ▷ Initialization
2: Estimate $\alpha_t \in [0,1]$ by solving the quadratic equation

$$L_t\alpha_t^2 = \Gamma_{t-1}(1-\alpha_t), \tag{2.50}$$

and set $\Gamma_t \leftarrow \Gamma_{t-1}(1-\alpha_t)$, $\nu_t = \max\{\nu_{t-1}, \alpha_t/\Gamma_t\}$.
3: Choose stepsize parameters as

$$\theta_t = \tau_t = \frac{\rho\nu_t\Gamma_t}{\alpha_t},\ \rho_t = \frac{\rho\alpha_t}{\Gamma_t\nu_t},\ \text{and } \eta_t = \frac{\Gamma_t}{\alpha_t} + \chi\theta_t M_t^2, \tag{2.51}$$

and calculate iterates (2.1) – (2.3).
4: **if** $G(x_{t+1}^{ag}) - G(x_t^{md}) - \langle \nabla G(x_t^{md}), x_{t+1}^{ag} - x_t^{md}\rangle > \frac{L_t}{2}\|x_{t+1}^{ag} - x_t^{md}\|^2$ **then**
5: Set $L_t \leftarrow 2L_t$. Go to 2. ▷ Backtracking L_G
6: **else if** $\chi\|Kx_{t+1} - Kx_t\| > \chi M_t\|x_{t+1} - x_t\|$ **then**
7: Set $M_t \leftarrow 2M_t$. Go to 2. ▷ Backtracking $\|K\|$
8: **end if**
9: **return** L_t, M_t, Γ_t, ν_t, x_{t+1}, x_{t+1}^{ag}, τ_t, ρ_t, α_t
10: **end procedure**

A few remarks are in place for the above backtracking procedure. Firstly, steps 2 through 8 are the backtracking steps, which terminates only when the conditions in steps 4 and 6 are both satisfied. Clearly, in each call to the backtracking procedure, steps 4 and 6 will only be performed finitely many times, and the returned values L_t and M_t satisfies $L_{min} \leq L_t \leq 2L_G$ and $M_t \leq 2\|K\|$, respectively. Secondly, while $M_t \geq M_{t-1}$ and $\nu_t \geq \nu_{t-1}$, the value of L_t in step 9 is not necessarily greater than L_{t-1}. Finally, the multiplier for increasing or decreasing L_t and M_t is 2, which can be replaced by any number that is greater than 1.

The scheme of AADMM with backtracking is presented in Algorithm 3.

Algorithm 3 AADMM with backtracking

Choose $x_1 \in X$ and $w_1 \in \mathcal{W}$ such that $Bw_1 = Kx_1 + b$, $L_0 \geq L_{min} > 0$ and $M_0, \nu_0, \rho > 0$. Set $x_1^{ag} \leftarrow x_1$, $w_1^{ag} \leftarrow w_1$, $y_1^{ag} \leftarrow y_1 = 0$, $\Gamma_0 \leftarrow L_0$, $t \leftarrow 1$.
for $t = 1, \cdots, N-1$ **do**
(L_t, M_t, Γ_t, ν_t, x_{t+1}, x_{t+1}^{ag}, τ_t, ρ_t, α_t)←BACKTRACKING(L_{t-1}, M_{t-1}, Γ_{t-1}, ν_{t-1}, x_t, x_t^{ag}, L_{min})
Calculate iterates (2.4) – (2.7).
end for

We start by considering UCO problems with bounded feasible sets X and Y. Theorem 2.11 below summarizes the convergence properties of Algorithm 3 for solving bounded UCO problems.

THEOREM 2.10. *If we set $\nu_0 = -\infty$ and apply Algorithm 3 to the UCO problem (1.4) under assumption (2.28), then*

$$\begin{aligned} &G(x_{t+1}^{ag}) + F(Kx_{t+1}^{ag}) - f^* \\ \leq\ &\frac{4L_G D_X^2}{t^2} + \frac{4L_G}{L_{min}(t-1)}\left[6\chi\rho\max\{4M_0^2, \|K\|^2\}D_X^2 + (1-\chi)\rho D_{X,K}^2 + \frac{D_Y^2}{\rho}\right]. \end{aligned} \tag{2.52}$$

In particular, if $\rho = \frac{D_Y}{\sqrt{6}\chi \max\{2M_0, \|K\|\}D_X + (1+\chi)D_{X,K}}$, *then*

$$\begin{aligned} & G(x_{t+1}^{ag}) + F(Kx_{t+1}^{ag}) - f^* \\ \le & \frac{4L_G D_X^2}{t^2} + \frac{4L_G}{L_{min}(t-1)}\left[\sqrt{6}\chi \max\{2M_0, \|K\|\}D_X D_Y + (1-\chi)D_{X,K}D_Y\right]. \end{aligned} \tag{2.53}$$

Proof. As discussed after Procedure 1, we have

$$L_{min} \le L_t \le 2L_G \text{ and } 0 \le M_t \le 2\|K\|. \tag{2.54}$$

We can now estimate the bounds of α_t *and* Γ_t. *By (2.12) we have* $1/\Gamma_t = 1/\Gamma_{t-1} + \alpha_t/\Gamma_t$, *hence*

$$\sqrt{\frac{1}{\Gamma_t}} - \sqrt{\frac{1}{\Gamma_{t-1}}} = \frac{1/\Gamma_t - 1/\Gamma_{t-1}}{\sqrt{1/\Gamma_t} + \sqrt{1/\Gamma_{t-1}}} = \frac{\alpha_t/\Gamma_t}{\sqrt{1/\Gamma_t} + \sqrt{1/\Gamma_{t-1}}}.$$

Observing from equations (2.12), (2.50) and (2.54) that

$$1/(2L_G) \le \alpha_t^2/\Gamma_t \le 1/L_{min}, \tag{2.55}$$

we have

$$\begin{aligned} \sqrt{\frac{1}{\Gamma_t}} - \sqrt{\frac{1}{\Gamma_{t-1}}} \ge \frac{\alpha_t/\Gamma_t}{2\sqrt{1/\Gamma_t}} = \frac{\alpha_t}{2\sqrt{\Gamma_t}} \ge \frac{1}{2\sqrt{2L_G}}, \text{ and} \\ \sqrt{\frac{1}{\Gamma_t}} - \sqrt{\frac{1}{\Gamma_{t-1}}} \le \frac{\sqrt{1/(L_{min}\Gamma_t)}}{\sqrt{1/\Gamma_t} + \sqrt{1/\Gamma_{t-1}}} \le \frac{1}{\sqrt{L_{min}}}. \end{aligned}$$

Therefore, by induction we conclude that

$$\frac{t}{2\sqrt{2L_G}} \le \sqrt{\frac{1}{\Gamma_t}} \le \frac{t}{\sqrt{L_{min}}} + \frac{1}{\sqrt{L_0}} \le \frac{t+1}{\sqrt{L_{min}}}, \text{ or } \frac{L_{min}}{(t+1)^2} \le \Gamma_t \le \frac{8L_G}{t^2}. \tag{2.56}$$

Now let us examine the RHS of (2.13) in Lemma 2.3. Without loss of generality, we assume that $2M_0 \le \|K\|$. *Indeed, if* $2M_0 > \|K\|$, *then* $M_t \equiv 2M_0$ *for all* $t \ge 1$. *Since* $\nu_{[t]}$ *and* $M_{[t]}$ *are monotonically increasing, by (2.51) and (2.27) in Lemma 2.4, we have*

$$\begin{aligned} \mathcal{B}_t(x^*, x_{[t+1]}, \eta_{[t]}) &\le \frac{1+\chi\nu_t\rho M_t^2}{2}(D_X^2 - \|x_{t+1} - x^*\|^2) \le \frac{1+\chi\nu_t\rho M_t^2}{2}D_X^2 \\ &\le \frac{1}{2}D_X^2 + 2\chi\nu_t\rho\|K\|^2 D_X^2, \\ \mathcal{B}_t(y, y_{[t+1]}, \rho_{[t]}^{-1}) &= \frac{\nu_t}{2\rho}(\|y_1 - y\|^2 - \|y_{t+1} - y\|^2) \le \frac{\nu_t}{2\rho}D_Y^2, \ \forall y \in \mathcal{Y}, \\ \mathcal{B}_t(w^*, w_{[t+1]}, \theta_{[t]}) &= \frac{\nu_t\rho}{2}(\|w_1 - w^*\|^2 - \|w_{t+1} - w^*\|^2) \le \frac{\nu_t\rho}{2}D_{X,K}^2. \end{aligned}$$

On the other hand, by (2.26) in Lemma 2.4 we have

$$\begin{aligned} -\chi\mathcal{B}_t(Kx^*, Kx_{[t+1]}, \theta_{[t]}) \le & -\frac{\chi\nu_1\rho}{2}(\|Kx_1 - Kx^*\|^2 - \|Kx_{t+1} - Kx^*\|^2) \le \frac{\chi\nu_1\rho}{2}\|K\|^2\|x_{t+1} - x^*\|^2 \\ & \le \frac{\chi\nu_t\rho}{2}\|K\|^2 D_X^2. \end{aligned}$$

Applying the above calculations on $\mathcal{B}(\cdot,\cdot,\cdot)$ *to Lemma 2.3, we have*

$$\begin{aligned} & \frac{1}{\Gamma_t}Q(w^*, x^*, y; z_{t+1}^{ag}) \le \frac{D_X^2}{2} + \frac{\nu_t\rho}{2}D_{X,K}^2 + \frac{5\chi\nu_t\rho}{2}\|K\|^2 D_X^2 + \frac{\nu_t}{2\rho}D_Y^2 \\ \le & \frac{D_X^2}{2} + 3\chi\nu_t\rho\|K\|^2 D_X^2 + \frac{(1-\chi)\nu_t\rho}{2}D_{X,K}^2 + \frac{\nu_t}{2\rho}D_Y^2. \end{aligned}$$

Observe that by (2.55) *and* (2.56), $\alpha_t/\Gamma_t \le (t+1)/L_{min}$, *and that*

$$\nu_t \le \max_{i=1,\dots t} \alpha_i/\Gamma_i \le (t+1)/L_{min}. \tag{2.57}$$

Using the previous two inequalities and (2.56), *we have*

$$\begin{aligned}\bar{g}_Y(z_{t+1}^{ag}) &\le \frac{4L_G D_X^2}{t^2} + \frac{24\chi\rho L_G(t+1)}{t^2 L_{min}}\|K\|^2 D_X^2 + \frac{4(1-\chi)\rho L_G(t+1)}{t^2 L_{min}} D_{X,K}^2 + \frac{4L_G(t+1)}{t^2 L_{min}\rho} D_Y^2 \\ &\le \frac{4L_G D_X^2}{t^2} + \frac{24\chi\rho L_G}{L_{min}(t-1)}\|K\|^2 D_X^2 + \frac{4L_G}{L_{min}\rho(t-1)} D_Y^2.\end{aligned}$$

The above inequality, in view of Proposition 2.2, then implies (2.52) *and* (2.53). ∎

For AECCO problems when both X and Y are bounded, we can also apply Algorithm 3 with $\chi = 0$, as long as the maximum number of iterations N is given. Theorem 2.11 below describes the convergence properties of AL-ADMM with backtracking for solving general AECCO problems.

THEOREM 2.11. *If we choose* $\chi = 0$, $\rho = 1$, *and* $\nu_0 = N/L_{min}$ *in Algorithm 3, then*

$$G(x_N^{ag}) + F(w_N^{ag}) - f^* \le \frac{4L_G D_{x^*}^2}{(N-1)^2} + \frac{4L_G D_{x^*,K}^2}{L_{min}(N-1)}, \ \textit{and} \tag{2.58}$$

$$\|Bw_N^{ag} - Kx_N^{ag} - b\| \le \frac{16\sqrt{L_G} D_{x^*}}{\sqrt{L_{min}}(N-1)^{3/2}} + \frac{16\sqrt{2}\sqrt{L_G} D_{x^*,K}}{\sqrt{L_{min}}(N-1)} + \frac{32 L_G D_{y^*}}{L_{min}(N-1)}. \tag{2.59}$$

Proof. In view of step 2 in Procedure 1, equation (2.57) *and the choice of* ν_0, *we can see that* $\nu_t \equiv N/L_{min}$. *By* (2.12), (2.50), (2.14) *and* (2.51), *we have*

$$\begin{aligned}\mathcal{B}_t(x^*, x_{[t+1]}, \eta_{[t]}) &= \frac{1}{2}(D_{x^*}^2 - \|x_{t+1} - x^*\|^2) \le \frac{1}{2} D_{x^*}^2, \\ \mathcal{B}_t(y, y_{[t+1]}, \rho_{[t]}^{-1}) &= \frac{N}{2L_{min}}(\|y_1 - y\|^2 - \|y_{t+1} - y\|^2), \ \forall y \in \mathcal{Y}, \\ \mathcal{B}_t(Bw^*, Bw_{[t+1]}, \theta_{[t]}) &= \frac{N}{2L_{min}}(\|Bw_1 - Bw^*\|^2 - \|B_{t+1} - Bw^*\|^2) \le \frac{N}{2L_{min}} D_{x^*,K}^2.\end{aligned}$$

Using the fact that $\tau_t \ge \rho_t$ *and* $\chi = 0$, *and applying the above calculations to Lemma 2.3, we have*

$$\frac{1}{\Gamma_t} Q(w^*, x^*, y; z_{t+1}^{ag}) \le \frac{1}{2} D_{x^*}^2 + \frac{N}{2L_{min}} D_{x^*,K}^2 + \frac{N}{2L_{min}}(\|y_1 - y\|^2 - \|y_{t+1} - y\|^2), \ \forall y \in \mathcal{Y}.$$

Similarly to the proof of Theorem 2.9, we have

$$\begin{aligned}&\|y_{t+1} - y^*\|^2 \le \frac{L_{min}}{N} D_{x^*}^2 + D_{x^*,K}^2 + D_{y^*}^2, \ \|y_1 - y_{t+1}\|^2 \le \frac{2L_{min}}{N} D_{x^*}^2 + 2D_{x^*,K}^2 + 4D_{y^*}^2, \ \textit{and} \\ &\frac{1}{\Gamma_t} Q(w^*, x^*, y; z_{t+1}^{ag}) + \frac{N}{L_{min}}\langle y_1 - y_{t+1}, y\rangle \le \frac{1}{2} D_{x^*}^2 + \frac{N}{2L_{min}} D_{x^*,K}^2, \ \forall y \in \mathcal{Y}.\end{aligned}$$

Setting $v_{t+1} = \Gamma_t N(y_1 - y_{t+1})/L_{min}$, $t = N-1$ *and applying* (2.56), *we have*

$$Q(w^*, x^*, y; z_N^{ag}) + \langle v_N, y\rangle \le \frac{4L_G D_{x^*}^2}{(N-1)^2} + \frac{4L_G D_{x^*,K}^2}{L_{min}(N-1)}, \tag{2.60}$$

$$\begin{aligned}\|v_N\| &\le \frac{8\sqrt{2}\sqrt{L_G N} D_{x^*}}{\sqrt{L_{min}}(N-1)^2} + \frac{8\sqrt{2} N\sqrt{L_G} D_{x^*,K}}{\sqrt{L_{min}}(N-1)^2} + \frac{16 N L_G D_{y^*}}{L_{min}(N-1)^2} \\ &\le \frac{16\sqrt{L_G} D_{x^*}}{\sqrt{L_{min}}(N-1)^{3/2}} + \frac{16\sqrt{2}\sqrt{L_G} D_{x^*,K}}{\sqrt{L_{min}}(N-1)} + \frac{32 L_G D_{y^*}}{L_{min}(N-1)}.\end{aligned} \tag{2.61}$$

These previous two relations together with Proposition 2.1 then imply (2.58) *and* (2.59). ∎

3. Numerical examples. In this section, we will present some preliminary numerical results of the proposed methods. The numerical experiments are carried out on overlapped LASSO, compressive sensing, and an application on partially parallel image reconstruction. All algorithms are implemented in MATLAB 2013b on a Dell Precision T1700 computer with 3.4 GHz Intel i7 processor.

3.1. Group LASSO with overlap. The goal of this section is to examine the effectiveness of the proposed methods for solving UCO problems with unbounded X. In this experiment, our problem of interest is the group LASSO model given by [27]

$$\min_{x\in\mathbb{R}^n} \sum_{i=1}^{m} (\langle a_i, x\rangle - f_i)^2 + \lambda \sum_{g\in\mathcal{G}} \|x_g\|, \tag{3.1}$$

where $\{(a_i, f_i)\}_{i=1}^m \subseteq \mathbb{R}^n \times \mathbb{R}$ is a group of datasets, x is the sparse feature to be extracted, and the structure of x is represented by group $\mathcal{G}$. In particular, $\mathcal{G} \subseteq 2^{\{1,\dots,n\}}$, and for any $g \subseteq \{1,\dots,n\}$, x_g is a vector that is constructed by components of x whose indices are in g, i.e., $x_g := (x_i)_{i\in g}$. The first term in (3.1) describes the fidelity of data observation, and the second term is the regularization term to enforce certain group sparsity. In particular, we assume that x is sparse in the group-wise fashion, i.e., for any $g \in \mathcal{G}$, x_g is sparse. Problem (3.1) can be formulated as a UCO problem (1.4) by defining the linear operator K as $Kx = \lambda(x_{g_1}^T, x_{g_2}^T, \dots, x_{g_l}^T)^T$, where $g_i \in \mathcal{G}$ and $\mathcal{G} = \{g_i\}_{i=1}^l$. Specially, if each g_i consists k elements, then (3.1) becomes

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|Ax - f\|^2 + \lambda\|Kx\|_{k,1}, \tag{3.2}$$

where $A = (a_1, \dots, a_m)^T$, $f = (f_1, \dots, f_m)^T$, and $\|\cdot\|_{k,1}$ is defined by $\|u\|_{k,1} := \sum_{i=1}^n \|(u^{(ki-k+1)}, \dots, u^{(ki)})^T\|$ for all $u \in \mathbb{R}^{kn}$, where $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^k$. Note that $F(\cdot) := \|\cdot\|_{k,1}$ is simple, so the solution of problem (1.2) can be obtained directly by examining the optimality condition, which is also known as soft-thresholding.

In this experiment, we generate the datasets $\{(a_i, f_i)\}_{i=1}^m$ by $f_i = \langle a_i, x_{true}\rangle + \varepsilon$, where $a_i \sim N(0, I_n)$, $\varepsilon \sim N(0, 0.01)$, and the true feature x_{true} is the n-vector form of a 64×64 two-dimensional signal whose support and intensities are shown in Figure 1. Within its support, the intensities of x_{true} are generated independently from standard normal distribution. We set $n = 4096$, $m = 2048$ and choose $\mathcal{G}$ to be all the 2×2 blocks in the 64×64 domain (so that $k = 4$), and apply L-ADMM, LP-ADMM, AL-ADMM and ALP-ADMM to solve (3.1) in which $\lambda = 1$. The parameters for AL-ADMM and ALP-ADMM are chosen as in Theorem 2.9, and N is set to 300. To have a fair comparison, we use the same Lipschitz constants $L_G = \lambda_{max}(A^TA) \approx 1.6 \times 10^4$, $\|K\| = 2$ and $\rho = 0.5$ for all algorithms without performing a backtracking. Both the primal objective function value $f(\tilde{x})$ and the feature extraction relative error $r(\tilde{x})$ at approximate solution $\tilde{x} \in \mathbb{R}$ versus CPU time are reported in Figure 1, where

$$r(\tilde{x}) := \frac{\|\tilde{x} - x_{true}\|}{\|x_{true}\|}. \tag{3.3}$$

From Figure 1 we can see that the performance of AL-ADMM and ALP-ADMM are almost the same, and both of them outperforms L-ADMM and LP-ADMM. This is consistent with our theoretical observations that AL-ADMM and ALP-ADMM have better rate of convergence (2.45) than ADMM (2.40).

3.2. Compressive sensing. In this subsection, we present the experimental results on the comparison of ADMM and AADMM for solving the following image reconstruction problem:

$$\min_{x\in X} \frac{1}{2}\|Ax - f\|^2 + \lambda\|Dx\|_{2,1}, \tag{3.4}$$

where x is the n-vector form of a two-dimensional image to be reconstructed, $\|Dx\|_{2,1}$ is the discrete form of the TV semi-norm, A is a given acquisition matrix (depending on the physics of the data acquisition), f represents the observed data, and $X := \{x \in \mathbb{R}^n : l_* \le x^{(i)} \le u_*, \forall i = 1, \dots, n\}$. Problem (3.4) is a special

case of UCO (1.4) with $\mathcal{W} = \mathbb{R}^{2n}$, $G(x) = \|Ax - b\|^2/2$, $F(w) = \|w\|_{2,1}$ and $K = \lambda D$. We assume that the finite difference operator D satisfies the periodic boundary condition, so that the problem in (2.2) with $\chi = 0$ can be solved easily by utilizing the Fourier transform (see [49]).

In our experiment, we consider two instances where the acquisition matrix $A \in R^{m\times n}$ is generated independently from a normal distribution $N(0, 1/\sqrt{m})$ and a Bernoulli distribution that takes equal probability for the values $1/\sqrt{m}$ and $-1/\sqrt{m}$ respectively. Both types of acquisition matrices are widely used in compressive sensing (see, e.g., [2]). For a given A, the measurements b are generated by $b = Ax_{true} + \varepsilon$, where x_{true} is a 64 by 64 Shepp-Logan phantom [47] with intensities in $[0, 1]$ (so $n = 4096$), and $\varepsilon \equiv N(0, 0.001I_n)$. We choose $m = 1229$ so that the compression ratio is about 30%, and set $\lambda = 10^{-3}$ in (1.6). Considering the range of intensities of x_{true}, we apply ALP-ADMM with parameters in Theorem 2.6 and LP-ADMM to solve (3.4) with bounded feasible set $X := \{x \in \mathbb{R}^n : 0 \le x^{(i)} \le 1, \forall i = 1, \dots, n\}$. It should be pointed that since $Y := \operatorname{dom} F^* = \{y \in \mathbb{R}^{2n} : \|y\|_{2,\infty} := \max_{i=1,\dots,n} \|(y^{2i-1}, y^{2i})^T\|_2 \le 1\}$, we have $D_X = D_Y = n$, which suggests that $\rho = 1/\|K\|$ may be a good choice for ρ. We also apply L-ADMM and AL-ADMM to solve (3.4), with $\chi = 1$ and $X = \mathbb{R}^n$. In this case we use the parameters in Theorem 2.9 with $N = 300$ for AL-ADMM. To have a fair comparison, we use the same constants $L_G = \lambda_{max}(A^T A)$ and $\|K\| = \lambda\sqrt{8}$ (see [9]) and $\rho = 1/\|K\|$ for all algorithms without performing backtracking. We report both the primal objective function value and the reconstruction relative error (3.3) versus CPU time in Figure 3.

It is evident from Figure 2 that AL-ADMM and ALP-ADMM outperforms L-ADMM and LP-ADMM in solving (3.1). This is consistent with our theoretical results in Corollaries 2.5, 2.6, 2.8 and 2.9. Moreover, it is interesting to observe that ALP-ADMM with box constrained X outperforms AL-ADMM with $X = \mathbb{R}^n$. This suggests that the knowledge of the ground truth is helpful in solving image reconstruction problems.

3.3. Partially parallel imaging. In this section, we compare the performance of AADMM with backtracking and Bregman operator splitting with variable stepsize (BOSVS) [12], which is a linearized ADMM method with backtracking, in reconstruction of magnetic resonance images from partially parallel imaging (PPI). In magnetic resonance PPI, a set of multi-channel k-space data is acquired simultaneously from radiofrequency (RF) coil arrays. The imaging is accelerated by sampling a reduced number of k-space samples. The image reconstruction problem can be modeled as

$$\min_{x\in X} \frac{1}{2} \sum_{j=1}^{n_{ch}} \|MFS_j x - f_j\|^2 + \lambda \|Dx\|_{2,1}, \tag{3.5}$$

where x is the vector form of a two-dimensional image to be reconstructed. In (3.5), n_{ch} is the number of MR sensors, $F \in \mathbb{C}^{n\times n}$ is a 2D discrete Fourier transform matrix, $S_j \in \mathbb{C}^{n\times n}$ is the sensitivity encoding map of the j-th sensor, and $M \in \mathbb{R}^{n\times n}$ describes the scanning pattern of MR sensors, and $X \subseteq \mathbb{C}^n$. In particular, S_j's and M are both diagonal matrices, and their diagonal vectors $\operatorname{diag} S_j \in \mathbb{R}^n$ and $\operatorname{diag} M \in \mathbb{R}^n$ are n-vector form of images that have the same dimension as the reconstructed image. In practice, $\operatorname{diag} S_j$ describes the sensitivity of the j-th sensor at each pixel, and $\operatorname{diag} M$ is a mask that takes value ones at the scanned pixels and zeros elsewhere. Figure 4 shows the two-dimensional image representations of $\{\operatorname{diag} S_j\}_{j=1}^{n_{ch}}$, x_{true} and $\operatorname{diag} M$. The PPI reconstruction problems are described in more details in [11]. It should be noted that (3.5) is a special case of (3.4), and that the percentage of nonzero elements in $\operatorname{diag} M$ describes the compression ratio of PPI scan. In view of the fact that $\|F\| = \sqrt{n}$, the Lipschitz constant L_G of (3.5) can be estimated by

$$L_G = \|\sum_{j=1}^{n_{ch}} S_j F^T M^2 F S_j\| \le n \|\sum_{j=1}^{n_{ch}} S_j\|^2 = n \|\sum_{j=1}^{n_{ch}} \operatorname{diag} S_j\|_\infty^2. \tag{3.6}$$

In this experiment, $n_{ch} = 8$, and the measurements $\{f_j\}_{j=1}^{n_{ch}}$ are generated by

$$f_j = M(FS_j x_{true} + \varepsilon_j^{re}/\sqrt{2} + \varepsilon_j^{im}/\sqrt{-2}),\ j = 1, \dots, n_{ch}$$

where the noises $\varepsilon_j^{re}, \varepsilon_j^{im}$ are independently generated from distribution $\mathcal{N}(0, 10^{-4}\sqrt{n}I_n)$. We generate four instances of experiments where the ground truth x_{true} are the human brain image (see Figure 4). The

Table 3.1: Data acquisition information in partially parallel image reconstruction.

Instance	Dimension of x_{true}	Sampling trajectory	Acquisition rate	L_G
1a	$n = 256 \times 256$	Cartesian mask	18%	3.34×10^5
1b	$n = 256 \times 256$	Pseudo random mask	24%	3.34×10^5
2a	$n = 512 \times 512$	Cartesian mask	18%	1.60×10^6
2b	$n = 512 \times 512$	Pseudo random mask	24%	1.60×10^6

information of the instances is listed in Table 3.1. In particular, instances 1a and 1b have Cartesian and pseudo-random k-space sampling trajectories respectively but share the same sensitivity map and ground truth, and so are instances 2a and 2b.

We first consider $X = \mathbb{C}^n$, and use AL-ADMM with backtracking to solve (3.5). We use the parameters in Theorem 2.11 with $N = 400$ in all PPI experiments. We also apply the BOSVS method in [12][2] to solve (3.5) with $X = \mathbb{C}^n$, which is a backtracking linesearch technique for L-ADMM with Barzilai-Borwein stepsize [3]. Furthermore, noticing that x_{true} is in bounded feasible set $X := \{x \in \mathbb{C}^n : |x^{(i)}| \leq 1, \forall i = 1, \ldots, n\}$, we also apply ALP-ADMM with backtracking to solve (3.5) with aforementioned bounded feasible set X. We set the parameters to $\lambda = 10^{-10}n$ in (3.5), and choose $L_0 = \|F\|^2 = n$, $L_{min} = L_G/10$, $M_0 = \|K\|/10 = \lambda\sqrt{8}/10$ for Algorithm 3 where L_G is listed in Table 3.1.

The performance of AL-ADMM, ALP-ADMM and BOSVS is shown in Figures 5 and 6, in terms of both the primal objective function value and relative error (3.3). It is evident that AL-ADMM and ALP-ADMM outperform BOSVS in terms of the decrement of both primal objective value and relative error to ground truth, especially in the case of using Cartesian sampling trajectory. Since the Cartesian sampling trajectory in our experiments collects less low-frequency data (the center part in the k-space) and has no randomness in sampling (see Figure 4), it makes harder to get a good reconstruction comparing with that of the pseudo-random sampling trajectory. Our experimental results indicates that in this case the AADMM is much more efficient than BOSVS in reconstruction. It is evident that AL-ADMM and ALP-ADMM outperform BOSVS in terms of the decrement of both primal objective value and relative error to ground truth. This observation is consistent with our theoretical result in Theorems 2.10 and 2.11.

4. Conclusion. We present in this paper the AADMM framework by incorporating a multi-step acceleration scheme into linearized ADMM. AADMM has better rates of convergence than linearized ADMM on solving a class of convex composite optimization with linear constraints, in terms of the Lipschitz constant of the smooth component. Moreover, AADMM can handle both bounded and unbounded feasible sets, as long as a saddle point exists. For the unbounded case, the estimation for the rate of convergence depends on the distance from initial point to the set of saddle points. We also propose a backtracking scheme to improve the practical performance of AADMM. Our preliminary numerical results show that AADMM is promising for solving large-scale convex composition optimization with linear constraints.

Acknowledgment. The authors would like to thank Invivo Philips, Gainesville, FL for providing the PPI brain scan datasets.

REFERENCES

[1] A. Auslender and M. Teboulle. Interior gradient and proximal methods for convex and conic optimization. *SIAM Journal on Optimization*, 16(3):697–725, 2006.

[2] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28(3):253–263, 2008.

[3] J. Barzilai and J. M. Borwein. Two-point step size gradient methods. *IMA Journal of Numerical Analysis*, 8(1):141–148, 1988.

[2]The BOSVS code is available at http://people.math.gatech.edu/~xye33/software/BOSVS.zip

[4] S. Becker, J. Bobin, and E. Candès. NESTA: a fast and accurate first-order method for sparse recovery. *SIAM Journal on Imaging Sciences*, 4(1):1–39, 2011.
[5] D. P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods*. Academic Press, 1982.
[6] D. P. Bertsekas. *Nonlinear programming*. Athena Scientific, 1999.
[7] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2011.
[8] R. S. Burachik, A. N. Iusem, and B. F. Svaiter. Enlargement of monotone operators with applications to variational inequalities. *Set-Valued Analysis*, 5(2):159–180, 1997.
[9] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical imaging and vision*, 20(1):89–97, 2004.
[10] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.
[11] Y. Chen, W. Hager, F. Huang, D. Phan, X. Ye, and W. Yin. Fast algorithms for image reconstruction with application to partially parallel MR imaging. *SIAM Journal on Imaging Sciences*, 5(1):90–118, 2012.
[12] Y. Chen, W. W. Hager, M. Yashtini, X. Ye, and H. Zhang. Bregman operator splitting with variable stepsize for total variation image reconstruction. *Computational Optimization and Applications*, 54(2):317–342, 2013.
[13] Y. Chen, G. Lan, and Y. Ouyang. Optimal primal-dual methods for a class of saddle point problems. *UCLA CAM report 13-31*, 2013.
[14] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Modeling & Simulation*, 4(4):1168–1200, 2005.
[15] A. d'Aspremont. Smooth optimization with approximate gradient. *SIAM Journal on Optimization*, 19(3):1171–1183, 2008.
[16] J. Douglas and H. Rachford. On the numerical solution of heat conduction problems in two and three space variables. *Transactions of the American mathematical Society*, 82(2):421–439, 1956.
[17] J. Eckstein and D. P. Bertsekas. On the douglasrachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(1-3):293–318, 1992.
[18] E. Esser, X. Zhang, and T. Chan. A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. *SIAM Journal on Imaging Sciences*, 3(4):1015–1046, 2010.
[19] D. Gabay. Applications of the method of multipliers to variational inequalities. In M. Fortin and R. Glowinski, editors, *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*, volume 15 of *Studies in Mathematics and Its Applications*, pages 299 – 331. Elsevier, 1983.
[20] D. Gabay and B. Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers & Mathematics with Applications*, 2(1):17–40, 1976.
[21] R. Glowinski and A. Marroco. Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de dirichlet non linéaires. *ESAIM: Mathematical Modelling and Numerical Analysis-Modélisation Mathématique et Analyse Numérique*, 9(R2):41–76, 1975.
[22] D. Goldfarb, S. Ma, and K. Scheinberg. Fast alternating linearization methods for minimizing the sum of two convex functions. *Mathematical Programming*, pages 1–34, 2010.
[23] T. Goldstein, B. ODonoghue, and S. Setzer. Fast alternating direction optimization methods. *CAM report*, pages 12–35, 2012.
[24] T. Goldstein and S. Osher. The split bregman method for l1-regularized problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343, 2009.
[25] B. He and X. Yuan. On the o(1/n) convergence rate of the douglas-rachford alternating direction method. *SIAM Journal on Numerical Analysis*, 50(2):700–709, 2012.
[26] M. R. Hestenes. Multiplier and gradient methods. *Journal of optimization theory and applications*, 4(5):303–320, 1969.
[27] L. Jacob, G. Obozinski, and J.-P. Vert. Group lasso with overlap and graph lasso. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 433–440. ACM, 2009.
[28] G. Lan. Bundle-level type methods uniformly optimal for smooth and non-smooth convex optimization. *Manuscript, Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL*, 2013.
[29] G. Lan, Z. Lu, and R. D. Monteiro. Primal-dual first-order methods with $\mathcal{O}(1/\varepsilon)$ iteration-complexity for cone programming. *Mathematical Programming*, 126(1):1–29, 2011.
[30] G. Lan and R. D. Monteiro. Iteration-complexity of first-order augmented lagrangian methods for convex programming. *Manuscript. School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta (May, 2009)*, 2009.
[31] G. Lan and R. D. Monteiro. Iteration-complexity of first-order penalty methods for convex programming. *Mathematical Programming*, pages 1–25, 2013.
[32] P.-L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis*, 16(6):964–979, 1979.
[33] Z.-Q. Luo. On the linear convergence of the alternating direction method of multipliers. *arXiv preprint arXiv:1208.3922*, 2012.
[34] R. D. Monteiro and B. F. Svaiter. Iteration-complexity of block-decomposition algorithms and the alternating direction method of multipliers. *SIAM Journal on Optimization*, 23(1):475–507, 2013.
[35] R. D. Monteiro and B. F. Svaiter. On the complexity of the hybrid proximal extragradient method for the iterates and the ergodic mean. *SIAM Journal on Optimization*, 20(6):2755–2787, 2010.
[36] R. D. Monteiro and B. F. Svaiter. Complexity of variants of Tseng's modified F-B splitting and Korpelevich's methods for hemivariational inequalities with applications to saddle-point and convex optimization problems. *SIAM Journal on*

Optimization, 21(4):1688–1720, 2011.

[37] J.-J. Moreau. Décomposition orthogonale dun espace hilbertien selon deux cônes mutuellement polaires.(french). *CR Acad. Sci. Paris*, 255:238–240, 1962.

[38] Y. Nesterov. Excessive gap technique in nonsmooth convex minimization. *SIAM Journal on Optimization*, 16(1):235–249, 2005.

[39] Y. E. Nesterov. A method for unconstrained convex minimization problem with the rate of convergence $O(1/k^2)$. *Doklady AN SSSR*, 269:543–547, 1983. translated as Soviet Math. Docl.

[40] Y. E. Nesterov. Smooth minimization of nonsmooth functions. *Mathematical Programming*, 103:127–152, 2005.

[41] J. Nocedal and S. J. Wright. *Numerical optimization*. Springer Science+ Business Media, 2006.

[42] H. Ouyang, N. He, L. Tran, and A. G. Gray. Stochastic alternating direction method of multipliers. In *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, pages 80–88, 2013.

[43] J. Pena. Nash equilibria computation via smoothing techniques. *Optima*, 78:12–13, 2008.

[44] M. J. D. Powell. A method for nonlinear constraints in minimization problems. In *Optimization (Sympos., Univ. Keele, Keele, 1968)*, pages 283–298. Academic Press, London, 1969.

[45] R. T. Rockafellar. *Convex analysis*. Princeton University Press (Princeton, NJ), 1970.

[46] R. T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization*, 14(5):877–898, 1976.

[47] L. A. Shepp and B. F. Logan. The fourier reconstruction of a head section. *Nuclear Science, IEEE Transactions on*, 21(3):21–43, 1974.

[48] P. Tseng. On accelerated proximal gradient methods for convex-concave optimization. *submitted to SIAM Journal on Optimization*, 2008.

[49] Y. Wang, J. Yang, W. Yin, and Y. Zhang. A new alternating minimization algorithm for total variation image reconstruction. *SIAM Journal on Imaging Sciences*, 1(3):248–272, 2008.

[50] X. Ye, Y. Chen, and F. Huang. Computational acceleration for MR image reconstruction in partially parallel imaging. *Medical Imaging, IEEE Transactions on*, 30(5):1055–1063, 2011.

[51] X. Ye, Y. Chen, W. Lin, and F. Huang. Fast MR image reconstruction for partially parallel imaging with arbitrary k-space trajectories. *IEEE Transactions on Medical Imaging*, 30(3):575–585, 2011.

Fig. 1: True feature x_{true} in the experiment of group LASSO with overlap. Left: the support of x_{true}. Right: the intensities of x_{true}.

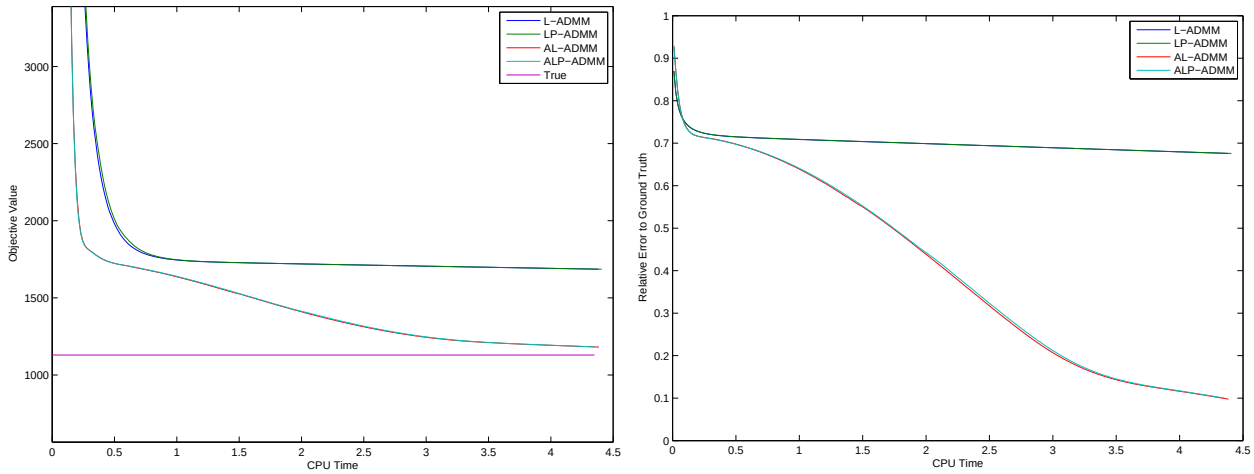

Fig. 2: Comparisons of AL-ADMM, ALP-ADMM, L-ADMM and LP-ADMM in group LASSO with overlap. Left: the objective function values $f(x_t^{ag})$ from AL-ADMM and ALP-ADMM, and $f(x_t)$ from L-ADMM and LP-ADMM vs. CPU time. The straight line at the bottom is $f(x_{true})$. Right: the relative errors $r(x_t^{ag})$ from AL-ADMM and ALP-ADMM and $r(x_t)$ from L-ADMM and LP-ADMM vs. CPU time.

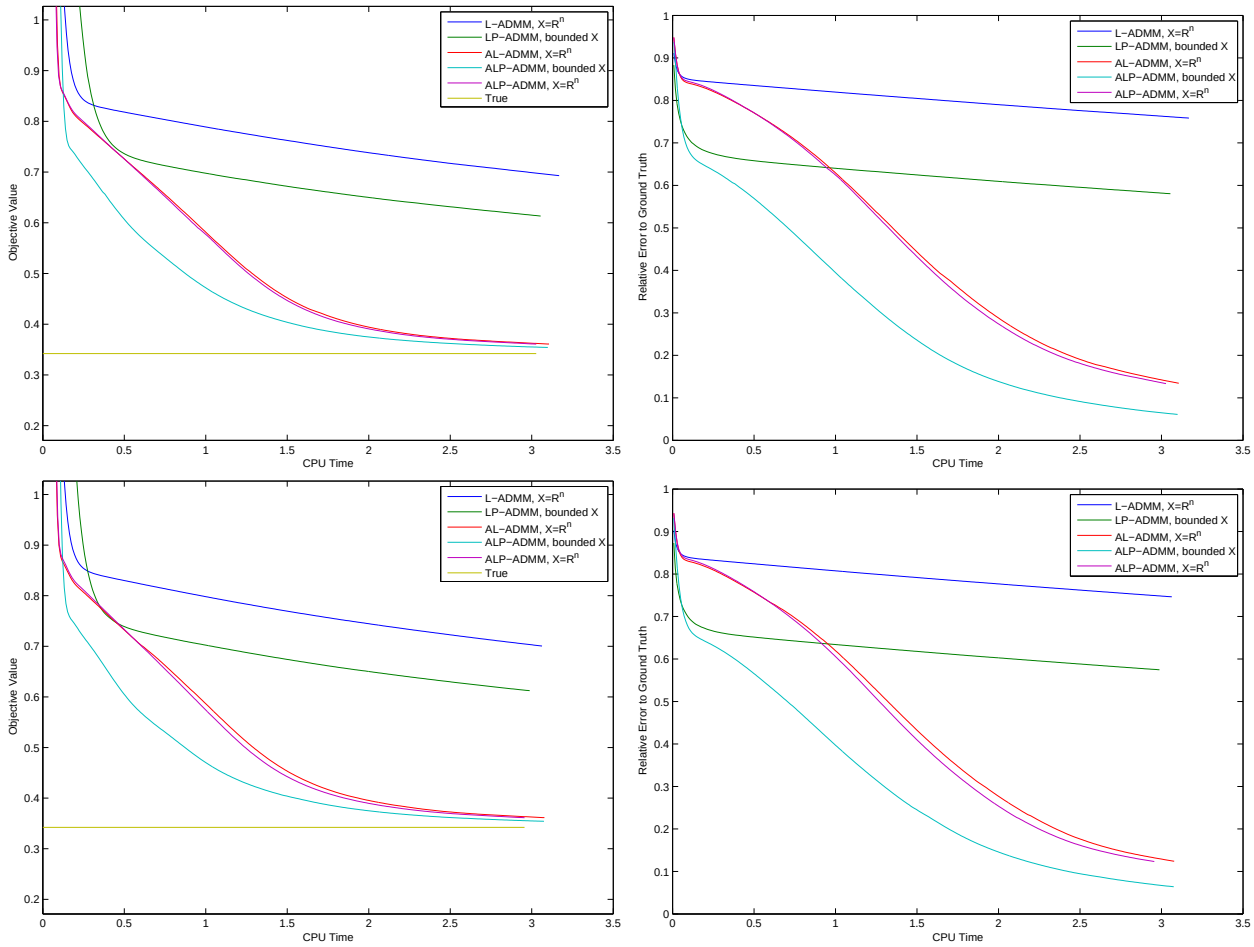

Fig. 3: Comparisons of AL-ADMM, ALP-ADMM, L-ADMM and LP-ADMM in image reconstruction. The top and bottom rows, respectively, show the performance of these algorithms on the "Gaussian" and "Bernoulli" instances. Left: the objective function values $f(x_t^{ag})$ from AL-ADMM and ALP-ADMM, and $f(x_t)$ from L-ADMM and LP-ADMM vs. CPU time. The straight line at the bottom is $f(x_{true})$. Right: the relative errors $r(x_t^{ag})$ from AL-ADMM and ALP-ADMM, and $r(x_t)$ in L-ADMM and LP-ADMM vs. CPU time.

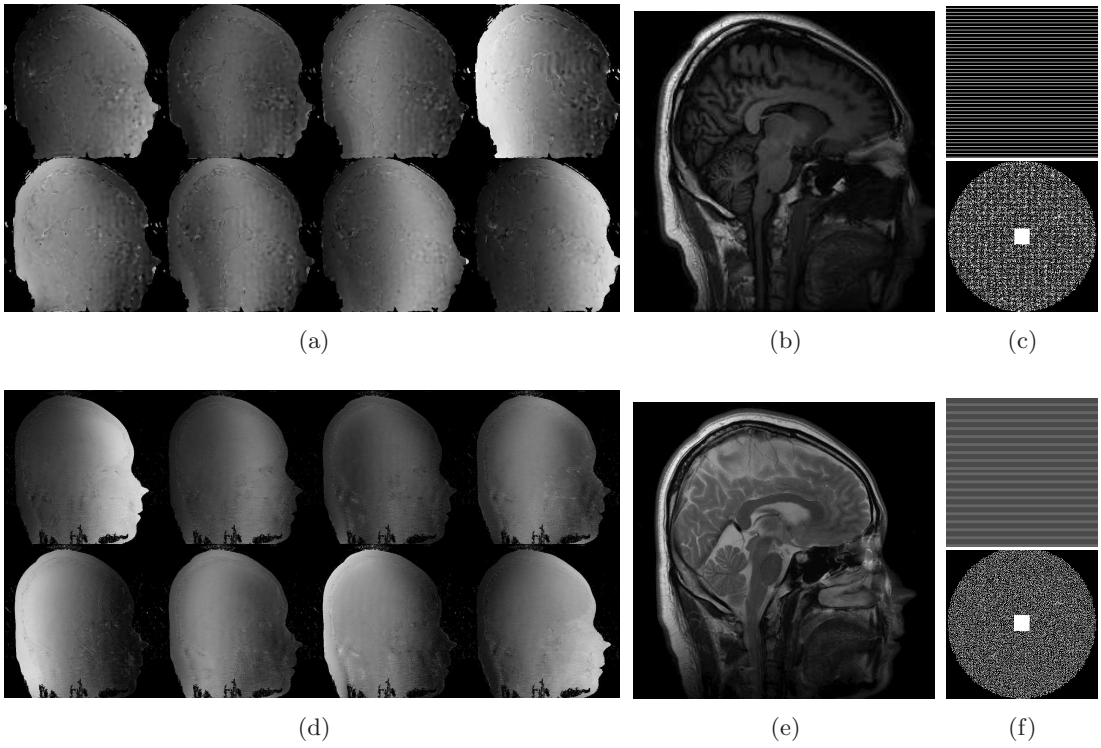

Fig. 4: Sensitivity map $\{\operatorname{diag} S_j\}_{j=1}^{8}$ (left), ground truth x_{true} (middle) and mask $\operatorname{diag} M$ (right) in partially parallel image reconstruction. (a): The sensitivity maps in instances 1a and 1b. (b): The ground truth in instances 1a and 1b. (c): The k-space sampling trajectory in instances 1a (top) and 1b (bottom). (d): The sensitivity maps in instances 2a and 2b. (e): The ground truth in instances 2a and 2b. (f): The k-space sampling trajectory in instances 2a (top) and 2b (bottom).

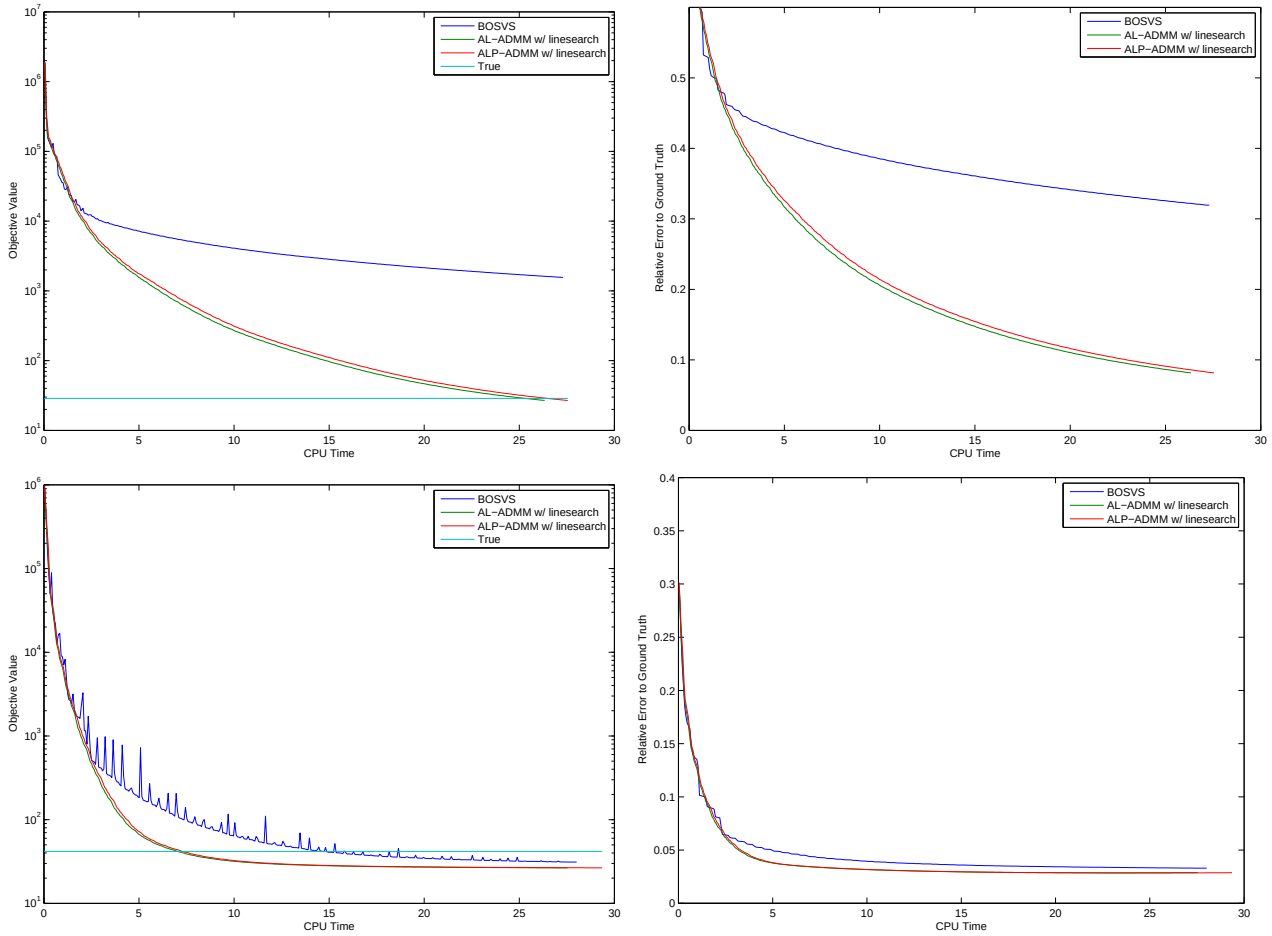

Fig. 5: Comparisons of AL-ADMM, ALP-ADMM and BOSVS in partially parallel image reconstruction. From top to bottom: performances of algorithms in instances 1a and 1b. Left: the objective function values vs. CPU time. The straight line at the bottom is $f(x_{true})$. Right: the relative errors vs. CPU time.

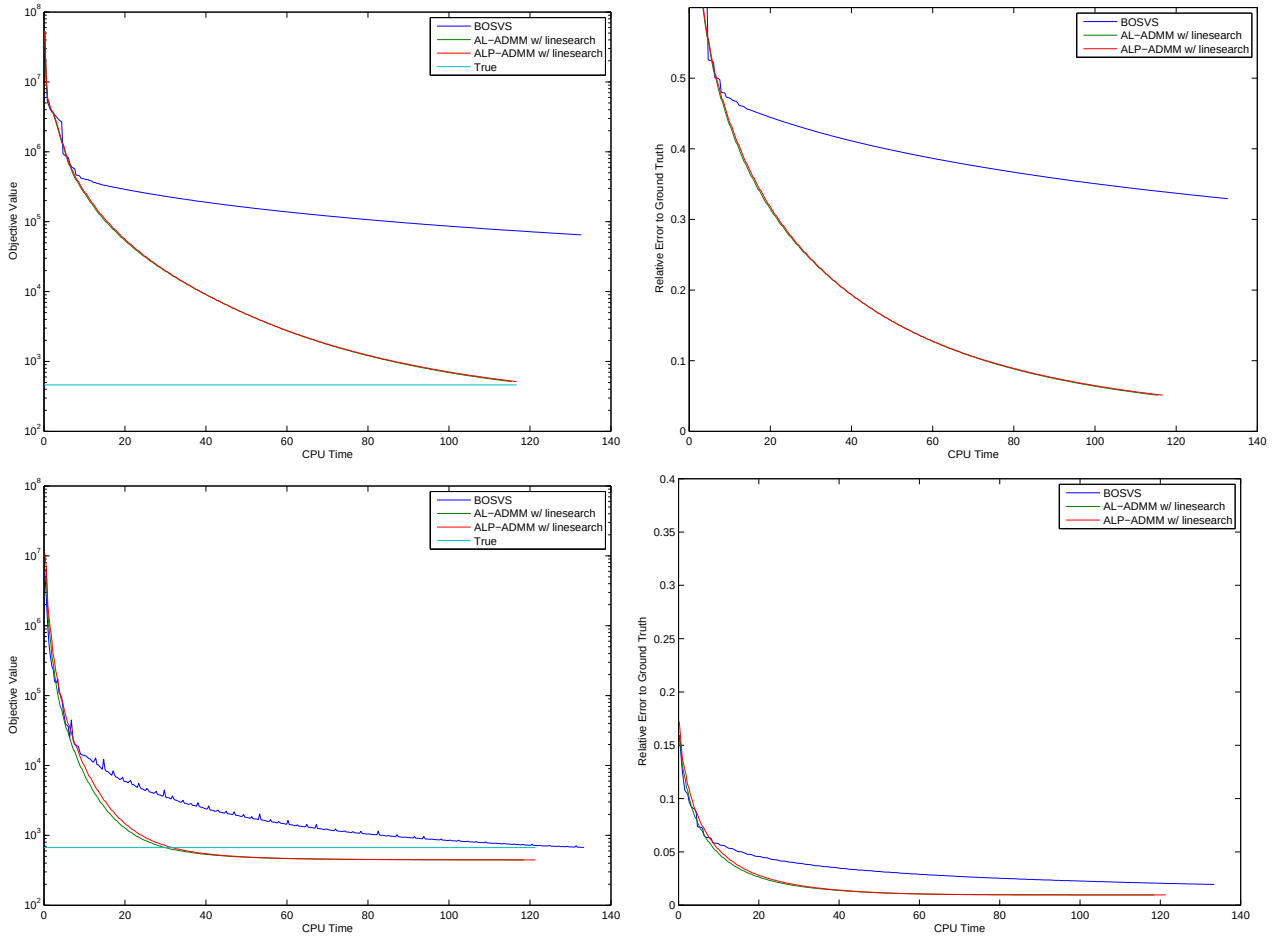

Fig. 5: Comparisons of AL-ADMM, ALP-ADMM and BOSVS in partially parallel image reconstruction (cont'd). From top to bottom: performances of algorithms in instances 2a and 2b. Left: the objective function values vs. CPU time. The straight line at the bottom is $f(x_{true})$. Right: the relative errors vs. CPU time.

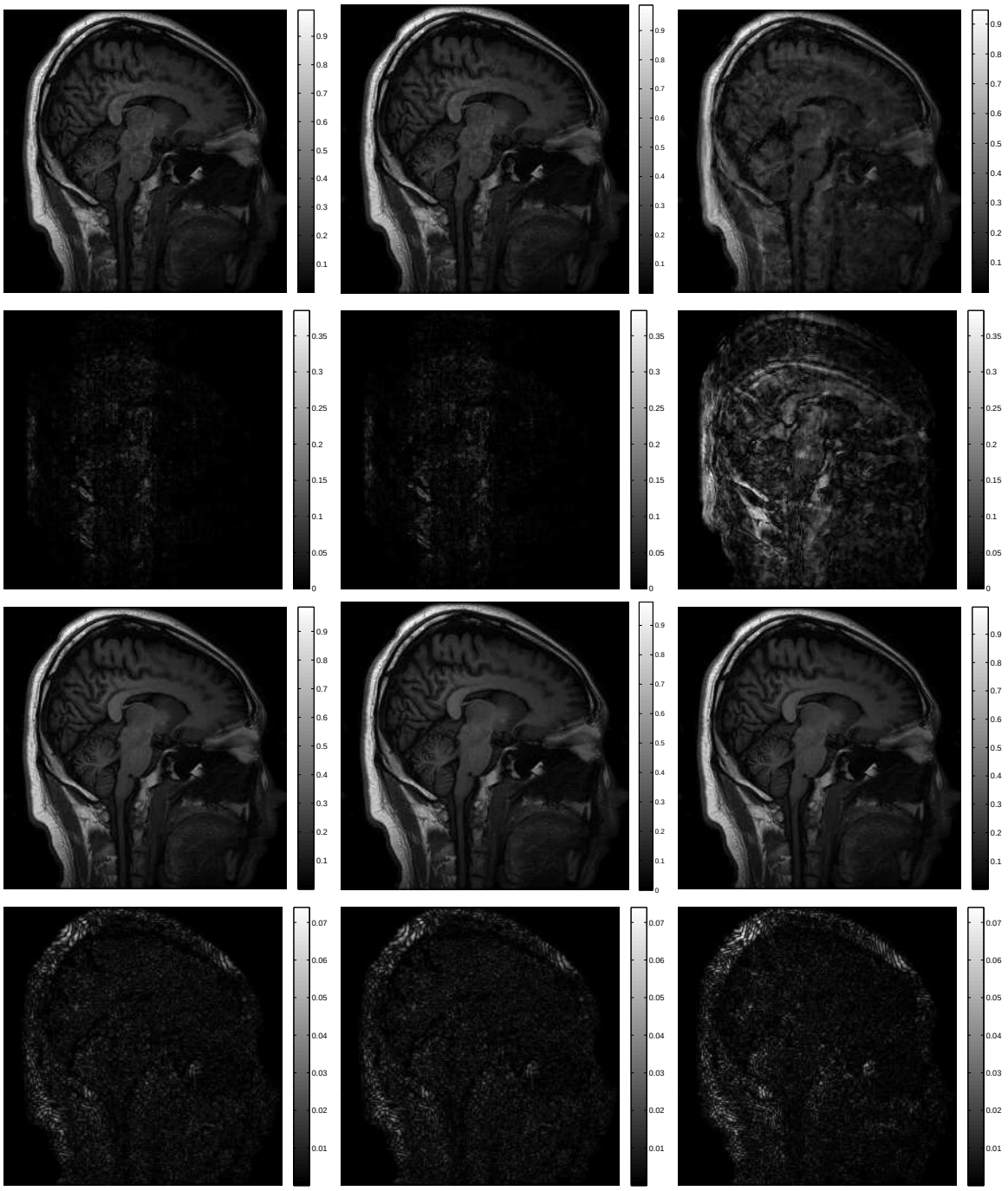

Fig. 6: Comparison of AL-ADMM, ALP-ADMM and BOSVS in partially parallel image reconstruction. From top to bottom: Reconstructed images and reconstruction errors in instances 1a and 1b, respectively. From left to right: AL-ADMM, ALP-ADMM and BOSVS.

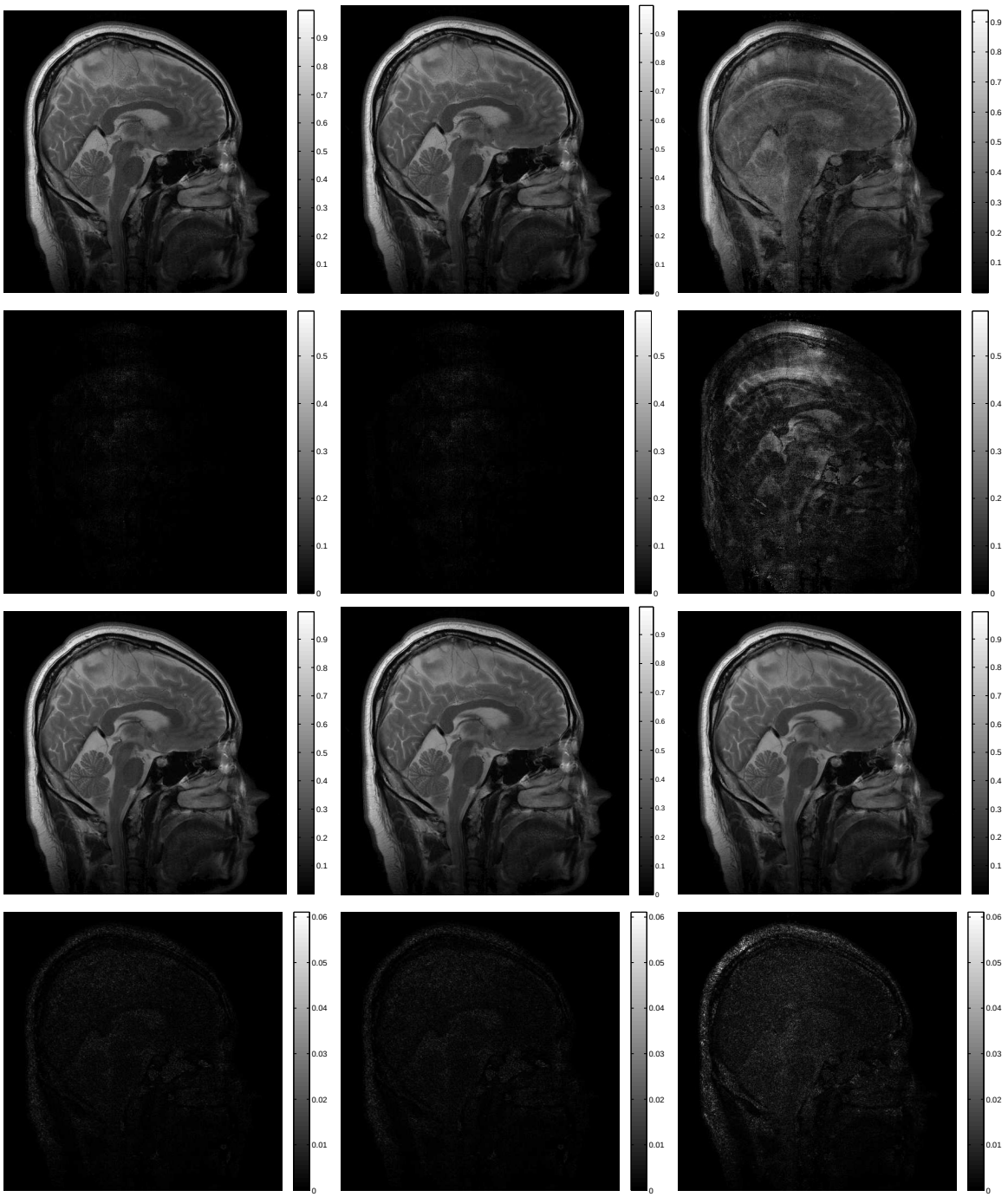

Fig. 6: Comparison of AL-ADMM, ALP-ADMM and BOSVS in partially parallel image reconstruction (cont'd). From top to bottom: Reconstructed images and reconstruction errors in instances 2a and 2b, respectively. From left to right: AL-ADMM, ALP-ADMM and BOSVS.